

UNIVERSIDADE FEDERAL DO PARANÁ

JOUGLAS ALVES TOMASCHITZ

**SEGMENTAÇÃO E ANÁLISE DE IMAGENS
DIGITAIS DE METÁFASES CROMOSSÔMICAS EM
BANDEAMENTO G**

CURITIBA

2013

UNIVERSIDADE FEDERAL DO PARANÁ

JOUGLAS ALVES TOMASCHITZ

**SEGMENTAÇÃO E ANÁLISE DE IMAGENS
DIGITAIS DE METÁFASES CROMOSSÔMICAS EM
BANDEAMENTO G**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Bioinformática, Setor de Educação Profissional e Tecnológica, Universidade Federal do Paraná, área de concentração Bioinformática.

Orientador: Prof. Dr. Luiz Antônio Pereira Neves

Co-orientador: Profa. Dra. Iris Hass

CURITIBA

2013

T655 Tomaschitz, Jouglas Alves
Segmentação e análise de imagens digitais de metáfases cromossômicas em bandeamento G / Jouglas Alves Tomaschitz. - Curitiba, 2013. - 96 f.: il., tabs, grafs.

Orientador: Prof º. Dr. Luiz Antônio Pereira Neves
Co-orientadora: Profa. Dra. Iris Hass
Dissertação (Mestrado) – Universidade Federal do Paraná, Setor de Educação Profissional e Tecnológica, Curso de Pós-Graduação em Bioinformática.
Inclui Bibliografia.

1. Cromossomos. 2. Processamento de imagens. 3. Bandeamento G. 4. Bioinformática. I. Neves, Luiz Antônio Pereira. II. Hass, Iris. III. Título. IV. Universidade Federal do Paraná.

CDD 572.87

Dedico à minha família... Mãe, Pai, Irmãos e Afilhada.
Em especial à minha esposa Joice

AGRADECIMENTOS

Agradeço os meus orientadores Prof. Dr. Luiz Antonio Pereira Neves e Prof.^a Dr.^a Iris Hass pela paciência, disponibilidade em ajudar e pelo conhecimento que me proporcionaram.

Ao Programa de Pós Graduação em Bioinformática da UFPR pela oportunidade.

Ao Professor Dr. Jacques Facon pelo incentivo para meu ingresso no Mestrado pela paciência e constante apoio durante todo o projeto.

Ao Professor Dr. Gilson Giraldi do LNCC o qual disponibilizou seu tempo para acrescentar ao projeto.

A MSc Terumi Kamada que me apoiou, incentivou e ajudou muitas vezes durante o projeto.

Aos meus colegas de trabalho Marcelo Jesus, Jonatan Pereira, Thays Xavier, que me apoiaram e ajudaram para que eu pudesse frequentar as aulas.

A todos os professores e colaboradores do PPG em Bioinformática.

Aos colegas da Bioinformática.

E especialmente:

- Ao meu pai e minha mãe, pelos ensinamentos de uma vida.
- Aos meus irmãos pelo constante apoio.
- A minha esposa Joice

“A maior sabedoria é conhecer a si mesmo”

Galileu

SUMÁRIO

SUMÁRIO	vi
LISTA DE FIGURAS	viii
LISTA DE tabelas	x
LISTA DE ABREVIATURAS SIGLAS	xi
RESUMO	xii
ABSTRACT	xiv
1 INTRODUÇÃO	1
1.1 OBJETIVOS	2
1.1.1 Objetivo geral	2
1.1.2 Objetivos específicos	2
1.2 JUSTIFICATIVAS	3
1.3 PROBLEMATIZAÇÃO	3
1.4 CONTRIBUIÇÕES ESPERADAS	3
1.5 ESTRUTURA DO TRABALHO	4
2 REVISÃO DE LITERATURA	5
2.1 Grupos cromossômicos	5
2.1.1 IDENTIFICAÇÃO DE CROMOSSOMOS	7
2.2 PROCESSAMENTO DIGITAL DE IMAGENS (PDI)	9
2.2.1 Definição da representação da imagem	11
2.2.2 Pré-processamento	12
2.2.2.1 Equalização de sub-imagens	12
2.2.1.2 ESCALA	13
2.2.3 Segmentação	14
2.2.3.1 Limiar de Ostu	16
2.2.3.2 Detecção de Bordas	16
2.2.3.3 Convex Hull	16
2.2.3.4 Limiarização de LAT (Local Average Thresholding)	17
2.2.3.5 Morfologia matemática - Abertura binária	18
2.2.3.6 Inversão da Imagem binária	18
2.2.3.7 Rotulação binária	19

2.2.3.8	Esqueletização Binária	19
2.2.3.9	LBP	21
2.2.4	Extração das características, Reconhecimento dos padrões e identificação	22
2.2.4.1	Fator compacidade	22
2.2.4.2	Análise de Clusters	23
2.2.4.3	Comparação DE HISTOGRAMAS utilizando o Chi-quadrado.....	23
2.2.5	APLICAÇÃO DE TÉCNICAS DE PROCESSAMENTO DE IMAGENS NA IDENTIFICAÇÃO AUTOMÁTICA DE CROMOSSOMOS	24
3	MATERIAIS E MÉTODOS	30
3.1	AMBIENTE HARDWARE.....	30
3.2	AMBIENTE SOFTWARE E LINGUAGEM DE PROGRAMAÇÃO	30
3.3	DESCRIÇÃO DA METODOLOGIA.....	30
3.3.1	Definição do modelo genérico para o processamento de imagens (PDI)	31
3.3.2	Definição da base de imagens.....	32
3.3.3	Métodos	32
3.3.4	Método 1	33
3.3.5	Método 2	36
3.3.6	Método 3	37
3.3.6.1	Modelo 3/1	42
3.3.6.2	Modelo 3/2	45
4	RESULTADOS	48
4.1	MÉTODO 1	48
4.2	MÉTODO 2	49
4.3	MÉTODO 3	50
4.3.1	Cálculo das Características	52
4.3.2	Cálculo de Características do Histograma LBP	62
4.4	ESCOLHA DO MÉTODO.....	65
5	DISCUSSÃO	66
6	CONCLUSÃO	72
7	REFERÊNCIAS.....	74

LISTA DE FIGURAS

FIGURA 1 – AGRUPAMENTOS CROMOSSÔMICOS.....	6
FIGURA 2 - REPRESENTAÇÃO DOS TONS DE CINZA	11
FIGURA 3 – PIRÂMIDE GAUSSIANA	13
FIGURA 4 - REPRESENTAÇÃO DO CONVEX HULL.....	17
FIGURA 5 – REPRESENTAÇÃO GRÁFICA.....	20
FIGURA 6 – APLICANDO O MÉTODO DE ESQUELETIZAÇÃO ZHANG-SUEN....	21
FIGURA 7 – APLICANDO AS 4 REGRAS.....	21
FIGURA 8 – OPERADOR LBP.....	22
FIGURA 9 – (a) CROMOSSOMO NA METÁFASE, (b) CARIÓTIPO DOS CROMOSSOMOS DA FIGURA 1 (A).....	25
FIGURA 10 – (a) DIVISÃO DO CROMOSSOMO EM SUB-REGIÕES, (b) RECONHECIMENTO DAS SUB-REGIÕES.....	25
FIGURA 11 – (a) METÁFASE DOS CROMOSSOMOS, (b) MASCARÁ OBTIDA A PARTIR DA LIMIAZIZAÇÃO.....	26
FIGURA 12 – IDEOGRAMA.....	26
FIGURA 13 – CLASSIFICAÇÃO.....	27
FIGURA 14– (a) IMAGENS DE CROMOSSOMOS EM BANDEAMENTO Q, (b) CARIÓTIPO.....	28
FIGURA 15 – ÁRVORE DE DECISÕES PARA RETIRADA DE SOBREPOSIÇÕES	28
FIGURA 16 – EIXO MÉDIO PRINCIPAL E PERFIL DE DENSIDADE.....	29
FIGURA 17 – MODELO GENÉRICO PARA O PROCESSAMENTO DE IMAGENS.....	32
FIGURA 18 – EXEMPLO DAS IMAGENS UTILIZADAS NA METODOLOGIA 1.....	33
FIGURA 19 – FLUXO DO MÉTODO 1.....	35
FIGURA 20 – RÓTULOS CRIADOS.....	36
FIGURA 21 – MODELO DO CROMOSSOMO 1.....	37
FIGURA 22 – VISÃO GERAL DO MODELO 3.....	38

FIGURA 23 – VISÃO GERAL DO MODELO 3 – SUDVISÃO.....	39
FIGURA 24 – IDENTIFICAÇÃO DO RÓTULO BINÁRIO.....	41
FIGURA 25 – VISÃO GERAL SUBDIVISÃO MODELO 3/1.....	43
FIGURA 26 – IMAGEM BINÁRIA.....	44
FIGURA 27 – IMAGEM DO ESQUELETO.....	44
FIGURA 28 – VISÃO GERAL SUBDIVISÃO MODELO 3/2.....	46
FIGURA 29 – IMAGEM SEGMENTADA.....	46
FIGURA 30 – IMAGEM LBP.....	47
FIGURA 31 – CRIAÇÃO DE RÓTULOS.....	49
FIGURA 32 – RECONHECIMENTO.....	50
FIGURA 33 – ETAPAS DE PROCESSO.....	51
FIGURA 34 – CROMOSSOMOS CLASSIFICADOS.....	56
FIGURA 35 – CROMOSSOMOS AGRUPADOS – GRUPO A, B.....	61
FIGURA 36 – CROMOSSOMOS AGRUPADOS – GRUPO C.....	61
FIGURA 37 – CROMOSSOMOS AGRUPADOS – GRUPOS D, E.....	61
FIGURA 38 – CROMOSSOMOS AGRUPADOS – GRUPOS F,G,C,G.....	61
FIGURA 39 - HISTOGRAMA LBP SUB-IMAGEM 3.....	62
FIGURA 40 - HISTOGRAMA LBP SUB-IMAGEM 24.....	62
FIGURA 41 - VISÃO GERAL DOS MÉTODOS.....	66

LISTA DE TABELAS

TABELA 1 – 4 VIZINHOS DE P.....	19
TABELA 2 – VIZINHANÇA DE 8 – CONDIÇÃO 1.....	19
TABELA 3 – VIZINHANÇA DE 8 – CONDIÇÃO 2.....	19
TABELA 4 – VIZINHANÇA DE 8 – CONDIÇÃO 3.....	20
TABELA 5 – VIZINHANÇA DE 8 – CONDIÇÃO 4.....	20
TABELA 6 – CARACTERÍSTICAS GERADAS.....	52
TABELA 7 – CLASSIFICAÇÃO.....	54
TABELA 8 – RESULTADOS.....	57
TABELA 9 – VALIDAÇÃO DE TODOS OS RESULTADOS	57
TABELA 10 – VALIDAÇÃO COM TODAS AS MEDIDAS.....	59
TABELA 11 – RESULTADO LBP.....	63
TABELA 12 – VALIDAÇÃO LBP.....	63
TABELA 13 – CARACTERÍSTICAS DOS MÉTODOS 1, 2 E 3.....	69

LISTA DE ABREVIATURAS SIGLAS

CCD – *Charge-Coupled Devices*

HSB – *Hue, Saturation and Brightness*

IBM – *International Business Machines*

IDE – *Integrated Development Environment*

LAT – *Local Average Threshold*

LBP – *Local Binary Pattern* (Padrão Local Binário)

LDA – *Análise do Discriminante Linear*

LNCC - *Laboratório Nacional de Computação Científica*

ML – *Maximum Likelihood*

NN – *Nearest Neighbor*

OPENCV – *Open Source Computer Vision*

PDI – *Processamento Digital de Imagens*

PCA – *Análise do Componente Principal*

RGB – *Red, Green, Blue*

SPSS - *Statistical Package for the Social Sciences*

RESUMO

A possibilidade de auxiliar diagnósticos genéticos, como na identificação de anomalias, com resultados em curto espaço de tempo, agiliza e direciona condutas de tratamento. O desenvolvimento do sistema semiautomático facilita o trabalho do geneticista. O processo de identificação dos cromossomos é prejudicado ao analisar imagens obtidas de fotomicrografia de baixa qualidade e resolução. Estudos citogenéticos automáticos de cromossomos metafásicos, por meio do Processamento Digital de Imagens (PDI), levam ao aperfeiçoamento e precisão no processo da especificação dos mesmos. Neste contexto, construir uma ferramenta computacional para facilitar o processo de identificação de cromossomos foi a motivação para esta pesquisa. A metodologia adotada é apresentada em três métodos diferentes, comparativos com o método genérico PDI de identificação do problema, aquisição da imagem, pré-processamento, segmentação, extração de características, reconhecimento, interpretação e resultado. Todas as imagens utilizadas foram de metáfases cromossômicas em bandejamento G. Para os Métodos 1 e 2 foram analisadas 30 imagens do roedor *Akodon montensis* do banco de imagens pertencentes ao acervo fotomicrográfico do Laboratório de Citogenética Animal, Departamento de Genética, Setor de Ciências Biológicas da UFPR. Para os Métodos 2 e 3 foram analisadas 30 imagens de cromossomos humanos (*Homo sapiens*), do banco de imagens Zooweb Karyotypes, Human Karyotypes for teaching, Wisconsin State Laboratory of Hygiene. O Método 3 foi subdividido em duas etapas, que apresentou resultados diferentes. Após diversos testes e estudos, verificou-se que o Método 3, mesmo com a divisão das etapas, foi o que teve melhor desempenho e resultados esperados no agrupamentos das classes cromossômicas A,B,C,D,E,F,G. O uso das técnicas de PDI mostrou que os resultados dependem da qualidade da imagem utilizada para se extrair as informações. No método 3/1 a análise de medidas de área obteve validação em todos os resultados com 76% de acerto. O método 3/2 obteve o reconhecimento do método LBP, histograma e comparação chi quadrado, com taxa de reconhecimento em 55%. Os resultados do histograma LBP trouxeram um nível de reconhecimento aceitável, porém a caracterização por medidas teve maior reconhecimento.

Palavras-chave: Segmentação Imagem, Bandeamento G, Metáfase, Cromossomos, Processamento de Imagens Digitais.

ABSTRACT

The possibility of assist in genetic diagnostics, such as in identifying anomalies with the results in a short time, accelerates and guide behavior of treatment. The development of semi-automatic system facilitates the work of geneticist. The process of identification of chromosomes is hindered when analyzing images obtained photomicrographs of low quality and low resolution. Cytogenetic studies automated of metaphase chromosomes by means of Digital Image Processing (DIP), lead to improve and accuracy in the process of specification. In this context, build a computational tool to facilitate the process of identification of chromosomes was the motivation for this research. The methodology is presented in three different methods, comparative with the PDI generic method of problem identification, image acquisition, preprocessing, segmentation, feature extraction, recognition, interpretation and results. All images used were of metaphase chromosome in G banding. For methods 1 and 2 were analyzed 30 images of rodent *Akodon montensis* from image bank photomicrographic belonging to the collection of the Laboratory of Animal Cytogenetics, Department of Genetics, Department of Biological Sciences UFPR. For Methods 2 and 3 were analyzed 30 images of human chromosomes (*Homo sapiens*), from image bank Zooweb Karyotypes, Human Karyotypes for teaching, Wisconsin State Laboratory of Hygiene. Method 3 was divided into two stages, which showed different results. After various tests and studies, it was found that Method 3, even with the division of the steps was the one with better performance and outcomes in the chromosomal groups of classes A, B, C, D, E, F, G. The use of the techniques of PDI showed that the results depends on the quality of the image used to extract the information. In method 3/1 analysis of area measurements obtained for all validation results with 76% accuracy. Method 3/2 obtained recognition of the LBP method, histogram and comparison chi square, with recognition rate of 55%. The results of the LBP histogram brought an acceptable level of recognition, but the characterization measures had greater recognition.

Keywords: Image Segmentation, G banding, metaphase, chromosomes, Digital Image Processing.

1 INTRODUÇÃO

Nos últimos anos cresceu o interesse em dados moleculares de doenças genéticas por parte de médicos, pesquisadores e profissionais da área de saúde. A possibilidade de se auxiliar o geneticista na identificação de anomalias com resultados em curto espaço de tempo, agiliza e direciona condutas de tratamento.

Os estudos citogenéticos que se baseiam na análise dos cromossomos são métodos essenciais no diagnóstico das doenças humanas. O processo de cariotipagem em algumas instituições até pouco tempo atrás ainda era feito de forma manual, o geneticista analisa o conjunto de cromossomos e os ordena de acordo com medidas e critérios de classificação e nomenclatura (REGATEIRO, 2007). Obedecendo aos critérios de alinhamento, o geneticista recortava os cromossomos do papel fotográfico e os catalogava conforme classificação. Os objetivos dos processos de cariotipagem são evidenciar possíveis desvios, tanto na estrutura, como no número de cromossomos e também na ordem de evolução da espécie. Assim, os estudos cromossômicos permitem localizar polimorfismos e heteromorfismos de cromossomos autossômicos e de cromossomos sexuais, ocorrência de poliploidia e de cromossomos supernumerários (BERTOLOTTI, 2006). Detectar desvios na estrutura e na quantidade de cromossomos permite compreender os mecanismos responsáveis pela variação cariotípica existente.

Os avanços ocorridos nos refinamentos das técnicas de diagnósticos genéticos com bandeamento, por meio do processamento de imagens do cariótipo, possibilitam progressos no aperfeiçoamento de identificação das alterações cromossômicas, porém mesmo com o aumento destas análises, ainda é pequena a oferta de ferramentas computacionais automáticas que auxiliem o trabalho dos geneticistas (KURTZ, 2011).

O aperfeiçoamento e precisão dos diagnósticos genéticos, obtida por meio da tecnologia, trazem responsabilidades éticas e legais e influenciam diretamente na terapêutica, na estimativa do prognóstico clínico e no cálculo do risco de recorrência para descendentes de acordo com coeficiente de consanguinidade dos pacientes (REGATEIRO, 2007). Diagnósticos pré-

sintomáticos de doenças de manifestação tardia podem ajudar no prognóstico, assim como o erro técnico pode ser prejudicial (GARCIA; CHAMAS, 1996 e VALENZUELA; ANTELO; ADOLFO, 2010). O envolvimento de pacientes e famílias no conhecimento dos fatores genéticos, que prejudicam a saúde, exige uma abordagem multiprofissional e interdisciplinar. O diagnóstico definitivo envolve, além de consultas, exames complementares. Seu significado e alcance dos resultados faz parte do processo dos pilares da propedêutica (BRUNONI, 2013).

Nesse contexto, os estudos citogenéticos automáticos, por meio da segmentação e análise das imagens digitais de cromossomos metafásicos, são importantes por permitirem uma verificação cromossômica mais detalhada, com maior qualidade e rapidez no processo de especificação, uma vez que são avaliadas imagens processadas para reconhecimento de padrões únicos.

1.1 OBJETIVOS

1.1.1 OBJETIVO GERAL

Construir uma ferramenta computacional para facilitar o processo de identificação de cromossomos humanos.

1.1.2 OBJETIVOS ESPECÍFICOS

- Segmentar e analisar imagens digitais de cromossomos metafásicos tratados com a técnica de bandeamento G, através dos métodos de processamento de imagens.
- Analisar o desempenho dos resultados nos métodos de processamento de imagens.

1.2 JUSTIFICATIVAS

Existem ferramentas que analisam e montam kariogramas para cromossomos humanos, porém estas ferramentas são acopladas a software e microscópios, de alto custo, e que não permitem a utilização de imagens externas ao equipamento, já que existem microscópios que podem, de forma independente, apenas fotografar o cromossomo utilizando uma câmera CCD (*Charge-Coupled Devices*) acoplada ao microscópio. O Processamento Digital de Imagens (PDI) se apresenta como uma ferramenta promissora na área de caracterização de materiais, devido a sua versatilidade quando comparada as técnicas convencionais, independentemente do tipo de material a ser caracterizado. O sucesso de uma abordagem diagnóstica depende, em muitos casos, fundamentalmente de acesso a ferramentas de forma precoce e precisa. O prognóstico e a terapêutica estão diretamente relacionados ao diagnóstico definitivo. O projeto contribui para isso, além de ajudar na evolução do diagnóstico citogenético.

1.3 PROBLEMATIZAÇÃO

Um dos grandes problemas existentes na área da Bioinformática é a aquisição de imagens de cromossomos por meio de imagens obtidas de fotomicrografia, com baixa qualidade e resolução. Esta limitação impede, em muitos casos, a identificação e análise dos cromossomos e estudos citogenéticos subsequente.

1.4 CONTRIBUIÇÕES ESPERADAS

Auxiliar o geneticista no processo de cariotipagem a fim de melhorar a qualidade e rapidez no processo de especificação do cromossomo. Este projeto poderá auxiliar o geneticista na classificação dos cromossomos de forma automática, a ferramenta construída será de baixo custo computacional e

financeiro e poderá ser ampliada em projetos futuros, introduzindo outras metodologias e comparações estatísticas, podendo assim contribuir para o aumento de pesquisas na área de forma mais rápida e eficaz.

1.5 ESTRUTURA DO TRABALHO

Este trabalho é composto de 7 capítulos. No primeiro capítulo, apresenta-se uma abordagem geral com introdução, objetivos, justificativas, problematização, relevância e contribuições esperadas da pesquisa.

No segundo capítulo efetua-se a revisão da literatura, contextualizando a fundamentação teórica, relacionando os principais autores e principais metodologias de identificação de cromossomos, processamento digital de imagens e aplicação correlata a estas áreas.

O terceiro capítulo descreve os materiais utilizados no desenvolvimento do trabalho, bem como a metodologia e os métodos desenvolvidos e implementados.

O quarto capítulo exhibe os resultados após a implementação dos métodos, realizados de acordo com a metodologia discutida na etapa anterior.

No quinto capítulo, tem-se a discussão, exibindo a diferença e a comparação entre os métodos implementados.

O sexto capítulo traz as considerações finais, conclusões e as recomendações para futuros trabalhos nesta área de pesquisa. E por último estão as referências bibliográficas.

2 REVISÃO DE LITERATURA

Este capítulo apresenta o estudo bibliográfico desta pesquisa, definindo a fundamentação teórica analisada, assim como os principais autores e principais metodologias das áreas relacionadas. Esta revisão tanto atende os conceitos biológicos quanto os conceitos computacionais, por ser uma pesquisa interdisciplinar.

A análise genética começa na divisão da célula em um conjunto de cromossomos, neste processo são corados e fotografados por meio de um microscópio (XIONG, 2007). Os cromossomos são estruturas espécie-específicas que contêm DNA, diferem pelo seu tamanho, localização do centrômero e por possuírem um padrão característico de bandas. Por meio de técnicas de coloração especial, que coram seletivamente o DNA, cada par cromossômico é individualmente identificado; isto ocorre durante a mitose, na metáfase, quando estão condensados ao máximo e quando os genes não podem ser transcritos (BORGES - OSÓRIO e ROBINSON, 2002). Após este processo é feito o cariótipo, conjunto total de cromossomos, que são utilizados em estudos clínicos e de evolução cromossômica, evidenciando possíveis anomalias tanto na estrutura como no número de cromossomos na espécie estudada. O cariótipo é apresentado como um arranjo sistematizado a partir de uma fotomicrografia de um único Núcleo Celular, organizado em pares em ordem decrescente de tamanho, sendo que atualmente a fotomicrografia foi substituída pela captura de imagem (metáfases cromossômicas) digitalizada (STEDMAN, 1995). Operações de segmentação efetuadas de modo manual baseiam-se no reconhecimento visual do objeto, por meio da comparação de diferenças apresentadas pelos elementos interpretativos, principalmente cor, tonalidade e textura (DLUGOSZ, 2009).

2.1 GRUPOS CROMOSSÔMICOS

As imagens cromossômicas são adquiridas por meio de um microscópio eletrônico e impressas em papel fotográfico. Para a construção de um

cariograma os cromossomos são recortados e ordenados conforme morfologia e ordem decrescente de tamanho e padrões de bandas. Os cromossomos humanos formam 23 pares sendo 22 autossomos e um par sexual. Os pares cromossômicos são ordenados de 1 até 22 e são reunidos em 7 grupos (A, B, C, D, E, F, G), como mostra Figura 1. Os cromossomos de cada grupo caracterizam-se por tamanho, localização do centrômero e padrões de bandas (KASAHARA, 2001). Os cromossomos agrupados da seguinte forma:

Grupo A: Pares 1, 2 e 3

Grupo B: Pares 4 e 5

Grupo C: Pares 6 ao 12, mais o cromossomo X

Grupo D: Pares 13, 14 e 15

Grupo E: Pares 16, 17 e 18

Grupo F: Pares 19 e 20

Grupo G: Pares 21, 22 e Y

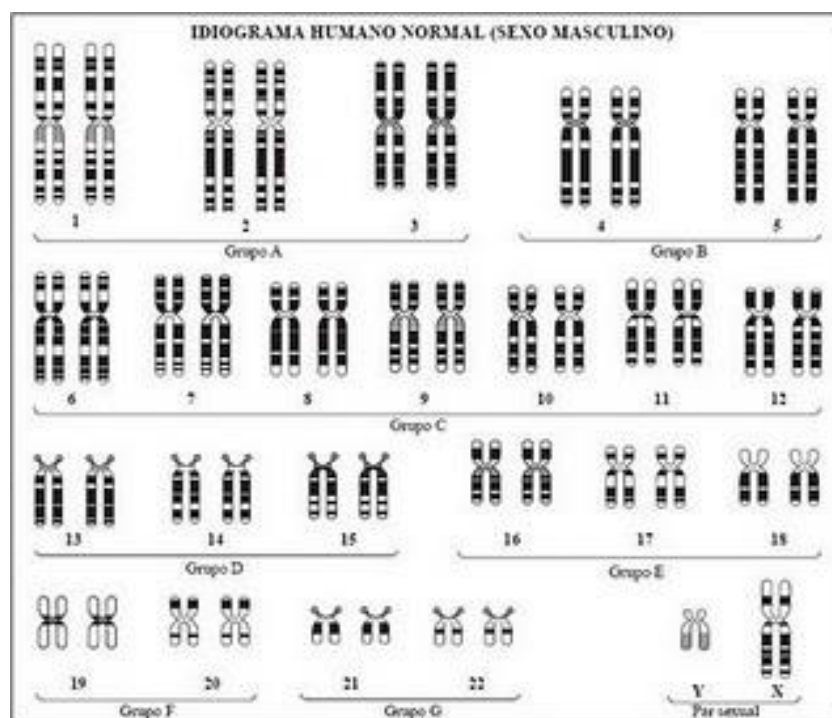


FIGURA 1 – AGRUPAMENTOS CROMOSSÔMICOS

2.1.1 IDENTIFICAÇÃO DE CROMOSSOMOS

O processamento de imagens, ferramenta utilizada para esta pesquisa, fornece muitas técnicas para melhoramento de imagens, com métodos para interpretação e utilização de diferentes etapas de processamento automáticos de segmentação, que procuram reconhecer áreas homogêneas na imagem, baseando-se em suas propriedades espectrais e espaciais (SPRING, 1998). A área está evoluindo continuamente, durante os últimos anos tem havido um aumento significativo no nível de interesse em morfologia matemática, redes neurais, processamento de imagens coloridas, compressão, reconhecimento de imagens e em sistemas de análise de imagens baseados em conhecimento (WOODS; GONZALEZ, 2002). O desenvolvimento da tecnologia tem ajudado várias áreas da ciência, mas ainda é pequena a oferta de sistemas automáticos que auxiliem o trabalho dos geneticistas na coleta de dados e geração do Cariótipo (KURTZ, 2008).

Na cariotipagem ou mapeamento do cariótipo, é feita a classificação e identificação dos cromossomos. A imagem dos cromossomos, após adquirida, passar por etapas de pré-processamento, segmentação, extração de características, reconhecimento, interpretação e resultado. A partir da imagem original, cria-se uma nova imagem ou máscara. A partir da imagem pré-processada, a segmentação separa cada uma das regiões de interesse, ou seja cada cromossomo, para armazená-los separadamente. Técnicas de reconhecimento de padrões e de ideogramas como modelos são utilizadas na tentativa de reconhecimento dos cromossomos. São comumente implementadas técnicas para o reconhecimento por tamanho e o reconhecimento por padrão de bandas ou bandeamento de cromossomos.

Pesquisas mostram que uma característica biométrica pode ser determinada a partir de outra. A comparação entre as características de um cromossomo com outro e a propriedade para parear cromossomos, pode ser assim considerada uma característica biométrica (SOMASUNDARAN; NIRMALA 2010). No final da década de 70 aparecem os primeiros trabalhos empregando técnicas de bandeamento que permitiram diferenciar, e da mesma forma,

evidenciar os pares cromossômicos em uma metáfase (PECCININI-SEALE et al., 1977a; PECCININI-SEALE et al., 1977b). Entende-se por Bandejamento de Cromossomos, Biblioteca Virtual em Saúde (DeCS, 2013):

“Coloração das bandas, ou segmentos cromossômicos, seguindo a identificação de cromossomos individuais ou partes de cromossomos. As aplicações incluem a determinação de rearranjos cromossômicos em síndromes de malformação e câncer, química de segmentos cromossômicos, alterações cromossômicas durante a evolução e, juntamente com estudos de hibridização, mapeamento cromossômico”.

O bandejamento cromossômico é caracterizado por coloração com marcação diferencial, tem importância para o pareamento cromossômico, pois cada par cromossômico apresenta um padrão distinto e característico de bandas. A identificação de alterações cromossômicas permite detectar pequenas variações estruturais, como deleções, inversões e translocações. A identificação de cromossomopatias, definidas como síndromes causadas por anomalias cromossômicas na espécie humana, pode ocorrer por alterações cromossômicas numéricas e estruturais. As alterações cromossômicas numéricas podem ser divididas em euploidias e aneuploidias, a Síndrome de Edwards é um exemplo de alteração no número de cromossomos. As alterações cromossômicas estruturais podem ser divididas em deleção, duplicação, inversão e translocação. Um exemplo de alterações cromossômicas estrutural é a Síndrome Cri-du-Chat ou Miado do gato. Para outras espécies animais e vegetais, as alterações numérica e/ou estruturais não representam somente a possibilidade de uma cromossomopatia, mas podem estar envolvidos em processos de evolução cromossômica (GUERRA, 1988).

Estudos da Localização da região do cromossomo diretamente afetada pode ajudar na terapêutica adequada para as síndromes e anomalias, além de elucidar a compreensão da evolução cariotípica entre as espécies (MARTINS, 2008).

As técnicas de bandejamento cromossômico fizeram parte da evolução da criptogenética (Casperson et al., 1970). O padrão de bandas de um cromossomo depende da técnica usada e do corante. As bandas Q foram as primeiras a serem

observadas, utiliza a microscopia de fluorescência, após coloração com quinacrina mostarda (ZECH, 1973; REGATEIRO, 2007). As bandas C são produzidas pela coloração da heterocromatina constitutiva com Giemsa, após desnaturação com hidróxido de bário (SUMNER, 1972). As bandas T estão localizadas nos telômeros e representam um subgrupo do Bandejamento R. As bandas N ou NOR englobam Região Organizadora de Nucléolos, ou seja, as "hastes" dos cromossomos acrocêntricos (REGATEIRO, 2007). Nas Bandas G ou GTG, consideradas padrões, os cromossomos são desproteinizados por ação da tripsina e, posteriormente são corados com Giemsa, de onde deriva o nome bandas G (Dretz & Shaw, 1971; SEABRIHT, 1971). Nesta técnica os cromossomos mostram um padrão de bandas claras e escuras, no qual as faixas escuras (Bandas G positiva) correspondem ao DNA rico em bases AT e poucos genes ativos; as bandas G claras (Bandas G negativas) têm DNA rico em bases GC, e apresentam muitos genes ativos.

2.2 PROCESSAMENTO DIGITAL DE IMAGENS (PDI)

Dentro da área de Processamento Digital de Imagens (PDI), relacionada à segmentação e análise, técnicas são utilizadas baseadas em descontinuidade e similaridade (NEVES, S. C. M.; PELAES, 2001). O método que avalia a descontinuidade considera a mudança abrupta dos valores de cinza e as técnicas baseadas em similaridades buscam dividir a imagem em regiões de interesse por meio das propriedades de distribuição dos pixels vizinhos (GONZALES; WINTZ, 1987).

WOODS e GONZALEZ (2002) afirmam que a seleção de critérios de similaridade depende do problema a ser solucionado e também do tipo de imagens disponíveis. Nas técnicas de processamento de imagens digitais, baseadas em similaridades busca-se dividir a imagem em regiões de interesse através das propriedades de distribuição dos pixels.

Os métodos mais comuns de segmentação por similaridade são caracterizados por algoritmos de limiarização (Tresholding), Aglomeração (Clustering), Crescimento de Regiões (Region Growing) e Separação e Junção

(Split & Merge). Na Limiarização (*Thresholding*), uma imagem de entrada $f(x, y)$ de N níveis de cinza, produz a saída de uma imagem $g(x, y)$, chamada de imagem limiarizada, cujo número de níveis de cinza é menor que N . Normalmente, $g(x, y)$ apresenta 2 (dois) níveis de cinza. Os pixels rotulados com 1 correspondem aos objetos, os pixels rotulados com 0 correspondem ao fundo, e T é um valor de tom de cinza predefinido denominado limiar. Neste caso tem-se a limiarização simples. A limiarização múltipla que admite dois ou mais limiares (T_1, T_2, T_3), com limites inferior e superior definidos níveis de cinza para segmentação, admitindo os níveis de cinza no intervalo (0; 255), (SEIXAS et al, 2008). O histograma de uma imagem é um conjunto de números indicando o percentual de pixels naquela imagem, que apresentam um determinado nível de cinza (NEVES, S. C. M.; PELAES, 2001). A visualização do histograma de uma imagem exibe uma indicação de qualidade relativa ao nível de contraste e ao seu brilho médio, quando a imagem é predominantemente clara ou escura. Quando o histograma possui um vale pode-se escolher o limiar como sendo o valor mínimo da curva do histograma, se houver mais que um vale tem-se o que a literatura chama de múltiplos limiares.

Existem várias maneiras de medir a similaridade entre dois padrões. Medidas de similaridade servem para comparar dois ou mais padrões em um processo de classificação ou medir a pertinência de um dado padrão a uma classe qualquer (MARTINS, 2005).

Outra técnica de segmentação por similaridade comumente utilizada é Aglomeração (*Clustering*), que se refere a tarefa de separar em classes de agrupamento de objetos em subconjuntos de acordo com certo critério; essa técnica atua sobre um espaço de atributos, que representa um espaço de dimensão elevada no qual cada ponto da imagem é representado por um vetor de atributos. (MARTINS, 2005). Extrair os atributos mais importantes em uma imagem evidencia as diferenças e similaridades entre os objetos. Algumas características são definidas por uma aparência visual na imagem usando RGB OU HSB (OLIVEIRA, 2008).

As técnicas de processamento digital de imagens baseadas em descontinuidades são definidas por uma operação orientada a vizinhança, um pixel da imagem filtrada depende não apenas do nível de cinza do pixel

correspondente na imagem original, mas também dos valores dos níveis de cinza dos pixels situados em sua vizinhança (QUEIROZ e GOMES, 2001). Os objetos que podem ser localizados por técnicas baseadas em descontinuidades são de três tipos: pontos, linhas e bordas. Operações lógicas e aritméticas orientadas a vizinhança utilizam o conceito de convolução com máscaras. A cada posição relativa da máscara sobre a imagem o pixel central da sub-imagem é substituído em uma matriz denominada imagem destino, pela soma dos produtos dos coeficientes com os níveis de cinza contidos na sub-região envolvida pela máscara (NEVES, S. C. M.; PELAES, 2001).

A imagem exibe uma informação, no processamento de imagens os conteúdos passam por formas de representação. Portanto, as etapas do processamento de imagens exibem o fluxo destas informações com objetivo definido por uma aplicação previamente estabelecida. O fluxo de forma geral vai, após a aquisição da imagem, do pré-processamento, segmentação, extração das características, reconhecimento dos padrões e identificação.

2.2.1 DEFINIÇÃO DA REPRESENTAÇÃO DA IMAGEM

Para representação da imagem, os pixels na imagem variam em tons de cinza representados pelos valores de 0 até 255. Onde o valor 0 na imagem representa o preto absoluto e o 255 o branco, como mostra Figura 2.

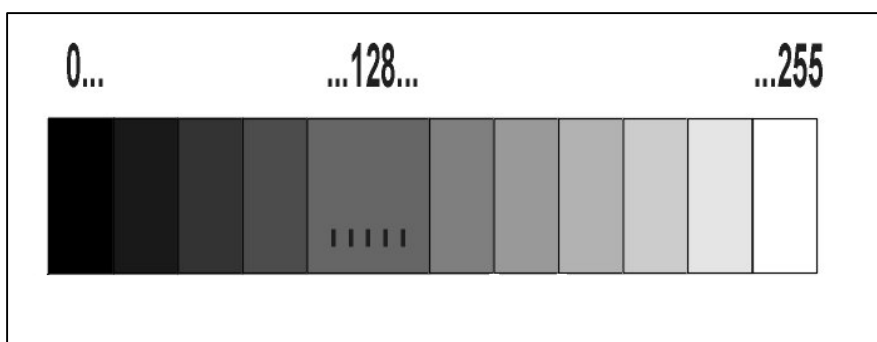


FIGURA 2 - REPRESENTAÇÃO DOS TONS DE CINZA

2.2.2 PRÉ-PROCESSAMENTO

O pré-processamento pode utilizar técnicas de realce de contraste, suavização ou até mesmo normalização de posicionamento do objeto a ser analisado.

2.2.2.1 EQUALIZAÇÃO DE SUB-IMAGENS

A fim de trazer uma melhoria no processamento, uma etapa importante é melhoria do contraste da imagem, a qual consiste na reorganização dos níveis de cinza da imagem cinza (WANG; CHEN; ZHANG, 1999).

O histograma nada mais é que uma distribuição probabilística dos níveis de cinza da imagem. Uma imagem com N pixels e coordenadas $X(i,j)$ e níveis de cinza $L=0$ até $L-1$ representando o nível mais escuro até o mais claro, $p(X_k)$ representa as distribuições das frequências, n_k é o número do total de pixels X_k . Como representado a seguir.

$$p(X_k) = \frac{n_k}{N}, \quad k = 0, 1, \dots, L - 1 \quad (1)$$

A função da distribuição cumulativa é definida como:

$$C(X_k) = \sum_{i=0}^k p(X_i) \quad (2)$$

A função transferência da equalização do histograma é definida como

$$f(X_k) = X_0 + (X_{L-1} - X_0)c(X_k), \quad k = 0, 1, \dots, L - 1 \quad (3)$$

A equalização por sub-imagens supõe uma imagem segmentada no nível de cinza $X = X_e$ em seções X_i e X_u de forma que $X = X_i$ e X_u . A sub-imagem X_i é composta pelos níveis de cinza $\{X_0, X_1, \dots, X_e - 1\}$ e X_u por $\{X_e, X_e + 1, \dots, L - 1\}$. Para cada sub-imagem é calculada a distribuição de frequências, distribuição cumulativa e transferências da equalização. Após as duas sub-imagens equalizadas são compostas em uma única. Representando Y como a imagem processada tem-se:

$$Y = \{Y(i, j)\} = f_L(X_L) \cup f_U(X_U) \quad (4)$$

Então

$$Y(i, j) = \begin{cases} 1. X_0 + (X_{e-1} - X_0)c_L(X), \\ 2. X_e + (X_{L-1} - X_e)c_U(X), \end{cases} \quad (5)$$

Calcula-se 1 se $X < X_e$, senão 2.

2.2.1.2 ESCALA

Neste método utiliza-se uma escala na imagem em 200%, que é suficiente para executar melhora a quantidade de informações presentes na imagem. O método de escala aplicado é o descrito no livro “Learning OpenCV” (BRADSKY, Gary; KAEHLER, 2008), que utiliza a Pirâmide Gaussiana Figura 3 (BRADSKY; PISAREVSKY; BOUGUET, 2006):

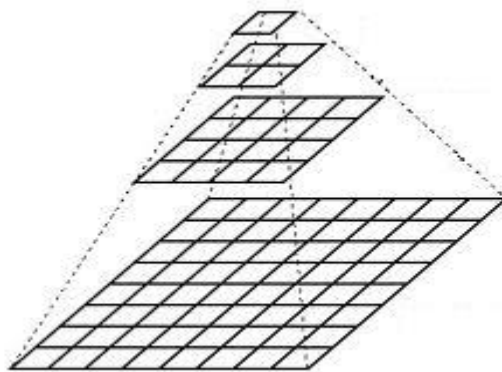


FIGURA 3 – PIRÂMIDE GAUSSIANA

- As Camadas são numeradas no sentido de baixo para cima, camada $(i+1)$. Demonstrada como G_{i+1} é menor que a camada $i(G_i)$.
- Para produzir uma camada $(i+1)$ na Pirâmide Gaussiana, faz-se:
 - Convolução de G_i com a máscara Gaussiana (Kernel Gaussian):

$$\frac{1}{16} \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}$$

MATRIZ 1 – MÁSCARA GAUSSIANA

- Remove-se cada par de números e da linha e coluna
- Nota-se que a imagem resultante é $\frac{1}{4}$ da área da imagem anterior. A interação deste processo na imagem inicial G_0 (imagem original) produz a pirâmide inteira.
- O procedimento anterior teve o objetivo de diminuir a imagem.
- Para aumentá-la executa-se o processo a seguir:
 - A imagem é aumentada ao dobro da imagem original em cada dimensão, as novas linha e colunas preencher com zeros (0).
 - Utiliza-se uma convolução com a mesma máscara apresentada anterior (multiplicar por 4) para aproximar os valores dos pixels em falta.

2.2.3 SEGMENTAÇÃO

Na segmentação a interpretação de um objeto considera-se não apenas a dimensão linear, como também a dimensão espacial. Para tanto, é necessário que o processo de segmentação seja moldado em função da resolução da imagem e da escala esperada para os objetos (Antunes, 2003). A rotulação utilizada normalmente é feita por regiões de interesse, onde estão os objetos separados, destacando a imagem de interesse com a cor branca e outra com a cor preta. Para a eliminação de ruídos utilizam-se filtros e para facilitar a visualização e a definição das regiões procuradas, aplica-se uma dilatação sobre a imagem, eliminando lacunas (KURTZ, 2008). Por sua vez, a erosão morfológica de uma imagem serve para a eliminação de detalhes que são irrelevantes em uma imagem. Portanto, baseando-se nas operações morfológicas de erosão e dilatação cria-se uma transformação, como a de remoção de ilhas ou pequenos objetos não relevantes (KURTZ, 2008). Na

identificação dos objetos pode-se fazer varreduras pela máscara para procurar pixels com valores específicos, quando encontrados os pixels vizinhos-de-4, estes são selecionados e armazenados (MARQUES FILHO; VIEIRA NETO, 1998). Após todas as regiões serem extraídas os objetos são retirados da imagem original utilizando-se a rotulação como referência. Por fim, armazenam-se as regiões de interesse da imagem separadamente, possibilitando o reconhecimento de padrões. Esta fase pode apresentar imagens de objetos irregulares, inclinados, ou sobrepostos, posteriormente algoritmos são desenvolvidos e adaptados para resolução destes problemas.

A limiarização, geralmente, primeiro passo no processamento de imagens para muitas aplicações, é baseada na similaridade do brilho dos objetos na imagem, com isso consegue-se fazer a segmentação do fundo da imagem e do objeto em estudo. Normalmente a maior parte dos métodos propostos selecionam limiarizações que dependem somente no histograma unidimensional da imagem, como no limiar de Otsu, que desenvolveu um método de seleção automática do nível de limiar, de modo a distinguir o melhor possível os níveis de cinza pertencentes ao fundo dos níveis de cinza componentes do objeto (OTSU, 1979); outras técnicas usam o LAT (Local Average Thresholding) e histogramas em 2D. Ao utilizar o histograma uni-dimensional em níveis de cinza, grande parte da informação contida na imagem pode ser perdida. Para se obter uma melhor segmentação, é preciso usar a maior quantidade de informação possível. Através da investigação da distribuição dos níveis de cinza / média local de cada classe (objeto ou fundo), a projeção ótima do histograma bidimensional pode ser derivada aplicando-se a Discriminante de Fisher. A projeção mostra ser, então, o histograma médio local. Em contraste com os grupos no histograma local de níveis de cinza, os grupos são mais compactados e mais separados. Limiarizar a projeção corresponde a separar o histograma em 2D com uma superfície planar ortogonal ao eixo médio local. Este método aprofunda-se no histograma 2D, podendo-se obter segmentações de imagens com baixo sinal em relação à ruído em tempo real (LI;Gong;Chen;1996).

2.2.3.1 LIMIAR DE OSTU

O limiar de otsu particiona os pixel em duas classes, C0 e C1, sendo que uma delas representam os objetos na imagem e outra representa o seu fundo. O seu limiar ótimo é calculado para cada imagem (OTSU,1979).

$$\sigma_w^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t) \quad (6)$$

Onde σ é a variância dentro da classe e ω probabilidade do particionamento das classes por um limiar t .

2.2.3.2 DETECÇÃO DE BORDAS

Na detecção de bordas, ocorre uma mudança abrupta dos níveis de cinza da imagem e do fundo da imagem. O detector de bordas de Canny, implementado nas bibliotecas do OpenCV, realça bordas existentes nas imagens (CANNY; 1986).

2.2.3.3 CONVEX HULL

É um conjunto de pontos os quais contornam objetos, criando rótulos. O convex hull cria um polígono ao redor dos objeto (Figura 4) e armazena as informações que ficam na área interna do polígono. (BERG; KERVELD; OVERMARS, 2000).

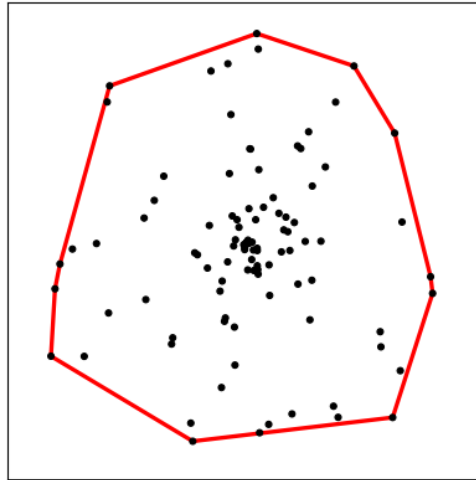


FIGURA 4 - REPRESENTAÇÃO DO CONVEX HULL

2.2.3.4 LIMIAÇÃO DE LAT (LOCAL AVERAGE THRESHOLDING)

Uma imagem que contem $N \times N$ pixels, seus níveis de cinza variam do mais escuro $L=0$ até o mais claro $L-1$. Para cada pixel o valor médio de vizinhança é calculado (LI; GONG, CHEN, 1996).

1. Fórmula 7 demonstra os grupos $N \times N$ percorridos, cada grupo é calculado

$$p_{ij} = \frac{f_{ij}}{N^2}$$

através da fórmula 8

$$, i, j = 0, \dots, L - 1. \quad (7)$$

- Passo 1 : Calcular a média local da imagem, $g(x,y)$, sobre uma janela quadrada $n \times n$ como mostrado na equação 8. Assumindo que $g(x,y)$ é a função do nível de cinza médio local, então:

$$g(x, y) = \frac{1}{n^2} \sum_{s=-n/2}^{n/2} \sum_{t=-n/2}^{n/2} f(x + s, y + t) \quad (8)$$

- Passo 2 : Obter o histograma local $\bar{h}(\cdot)$ da imagem para que se tenha $\bar{h}(j) = \{r_j \mid 0 \leq j < L\}$, onde r_j representa a média de um nível de cinza.
- Passo 3 : Escolher uma limiarização, T , aplicando uma abordagem de seleção de limiarização 1D, como o método de Otsu, Kapur, etc., ao histograma médio local, neste caso foi aplicado o Otsu.
- Passo 4 : Particionar a imagem em duas classes em um nível de cinza médio local T , então:

$$f_T(x, y) = \begin{cases} b_0, & \text{se } g(x, y) < T \\ b_1, & \text{se } g(x, y) \geq T \end{cases} \quad (9)$$

onde $0 \leq b_0, b_1 < L$, $f_T(x, y)$ é a função da limiarização média local (Local Average Thresholding – LAT).

2.2.3.5 MORFOLOGIA MATEMÁTICA - ABERTURA BINÁRIA

A abertura binária é uma operação originária da morfologia matemática onde é realizada uma erosão seguida de uma dilatação na imagem. Este processo é utilizado projeto a fim de fechar os “buracos” deixados pela limiarização, assim evita-se perder alguma informação da imagem (FACON, 1996).

2.2.3.6 INVERSÃO DA IMAGEM BINÁRIA

Inversão da imagem binária ou operação de NOT é o processo onde há inversão dos valores. A imagem base contém somente pixels pretos e brancos, os pixels pretos são convertidos em brancos e os brancos são convertidos em pretos.

2.2.3.7 ROTULAÇÃO BINÁRIA

A rotulação dos pixels em uma imagem binária utilizando um rótulo R, o algoritmo verifica os pixels que estão em vizinhança através do algoritmo 4-vizinhos e os armazena em um mesmo rótulo. Caso o pixel encontrado não esteja na mesma vizinhança é criado um novo rótulo. Todos os pixels da sequência P, P2..P8 devem ser pretos (SHAPIRO; STOCKMAN, 2002).

TABELA 1 – 4 VIZINHOS DE P

	P2	
P8	P	P4
	P6	

2.2.3.8 ESQUELETIZAÇÃO BINÁRIA

A esqueletização consiste em uma operação morfológica onde se obtém apenas uma representação compacta da imagem, geralmente em uma linha contínua. Para que um esqueleto seja obtido alguns requisitos precisam ser satisfeitos. (ZHANG, SUEN, 1984).

1. O número de conectividade é 1;

TABELA 2 – VIZINHANÇA DE 8 – CONDIÇÃO 1

P9	P2	P3
P8	P1	P4
P7	P6	P5

O número de conectividade é a transição de branco para preto que deve ser somente de uma vez, começando em P2 e terminando em P9, circundando o pixel central.

2. De P2 a P9 devem existir ao menos dois pixels vizinhos pretos, e não mais do que seis;

TABELA 3 – VIZINHANÇA DE 8 – CONDIÇÃO 2

P9	P2	P3
P8	P1	P4
P7	P6	P5

3. Ao menos um dos $I(i-1,j)$, $I(i,j+1)$ e $I(i+1, j)$ é um pixel branco;

TABELA 4 – VIZINHANÇA DE 8 – CONDIÇÃO 3

P9	P2	P3
P8	P1	P4
P7	P6	P5

4. Ao menos um dos $I(i,j+1)$, $I(i+1,j)$ e $I(i, j-1)$ é um pixel branco;

TABELA 5 – VIZINHANÇA DE 8 – CONDIÇÃO 4

P9	P2	P3
P8	P1	P4
P7	P6	P5

Representando graficamente o processo acima temos:

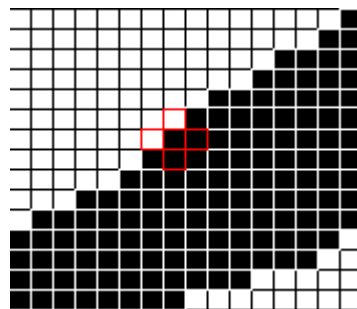


FIGURA 5 – REPRESENTAÇÃO GRÁFICA

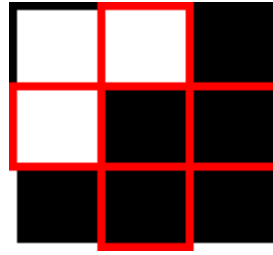


FIGURA 6 – APLICANDO O MÉTODO DE ESQUELETIZAÇÃO ZHANG-SUEN

Aplicando as 4 regras verifica-se que todas são atendidas então o pixel central é removido.

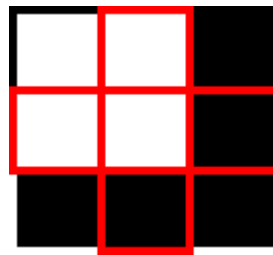


FIGURA 7 – APLICANDO AS 4 REGRAS

O algoritmo itera e caso não haja mais pixels a serem removidos a esquelização está completa.

2.2.3.9 LBP

O LBP é um operador para a análise de textura. Segundo a fórmula 10, P é a posição do pixel, R é raio escolhido na análise, $I_{p,R}$ pixel que está sendo verificado no momento e I_c é o pixel central (ZHAO; YONGSHENG; ZHANG, 2008).

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(I_{p,R} - I_c) 2^p \quad (10)$$

Para que o LBP seja criado, então, verifica-se $s(x)$, fórmula 11, caso o valor seja maior ou igual a zero, então o local recebe o valor 1; caso contrário receberá 0.

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (11)$$

Após a execução do algoritmo sobre a janela de raio R, o valor binário para o local é atribuído de forma horária, une-se os dígitos que representaram o número binário, que são transformados para decimal, como representado abaixo:

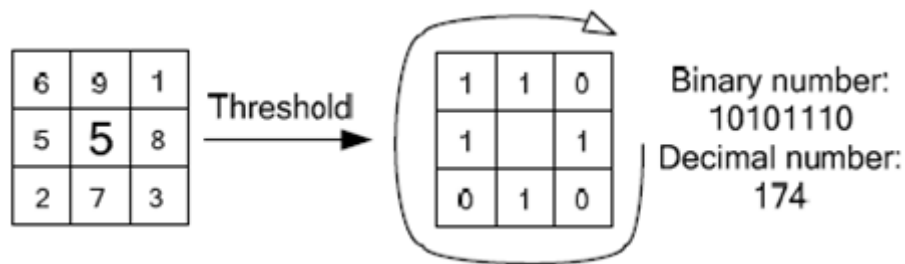


FIGURA 8 – OPERADOR LBP

2.2.4 EXTRAÇÃO DAS CARACTERÍSTICAS, RECONHECIMENTO DOS PADRÕES E IDENTIFICAÇÃO

As medidas utilizadas são medidas relativas porque as imagens pode se diferenciar em escala. Medidas como o tamanho, área, perímetro e altura são eficazes para a caracterização de padrões. A descrição quantitativa destas informações através da extração de características estruturais complementa o sentido de representação. Em seguida, com base na descrição, o reconhecimento associa um rótulo a cada objeto segmentado enquanto a interpretação associa um significado ao conjunto de objetos segmentados.

2.2.4.1 FATOR COMPACIDADE

Fator compacidade, o qual faz o cálculo de quanto a imagem é alongada ou arredondada. É uma boa medida para caracterização de formas de objetos.

$$Fc = \frac{p^2}{4 * \pi * S} \quad (12)$$

FONTE: FELIPE , NEVES (2008)

2.2.4.2 ANÁLISE DE CLUSTERS

Os clusteres hierárquicos utilizam matrizes de semelhanças e diferenças para suas descrições. As descrições dos graus de semelhança e diferenças são dadas por algoritmos de agrupamentos, neste caso foi utilizado o agrupamento por centroide, o qual utiliza o centro de massa para definir o centroide.

Os agrupamentos são calculados através de medidas de vizinhança neste caso foi utilizado a distância euclidiana ao quadrado. A distância da vizinhança é calculada pelos pontos X_{iv} e X_{jv} o menor índice resultante do cálculo apresenta maior vizinhança d_{ij}^2

$$d_{ij}^2 = \sum_{v=1}^p (X_{iv} - X_{jv})^2 \quad (13)$$

Carrega-se o Arquivo de Características, utiliza-se o software SPSS e analisa-se as similaridades dos clusteres.

2.2.4.3 COMPARAÇÃO DE HISTOGRAMAS UTILIZANDO O CHI-QUADRADO.

Para comparação de histogramas utiliza-se o coeficiente chi-quadrado, que analisa um valor de dispersão entre duas variáveis. Então na análise dos histogramas quanto mais um coeficiente se aproximar de zero, mais similaridade se tem. H^P representa o histograma da imagem a qual será comparada com todas as outras H^G .

$$\chi^2(H^P, H^G) = \sum_{i=0}^l \frac{(H_i^P - H_i^G)^2}{(H_i^P + H_i^G)} \quad (14)$$

2.2.5 APLICAÇÃO DE TÉCNICAS DE PROCESSAMENTO DE IMAGENS NA IDENTIFICAÇÃO AUTOMÁTICA DE CROMOSSOMOS

No trabalho de XIONG (2007), com o objetivo de eliminar ruídos, evitar perdas de informação e realçar as bordas de uma imagem utiliza uma transformada de Wavelet. A partir das imagens originais são criados vetores contendo sub-imagens e destes vetores são criados conjuntos de treino utilizando características do PCA e LDA e classificadores NN e ML, para a classificação do cromossomo como mostra Figura 9. A transformada de wavelet decompõe uma função definida no domínio do tempo em outra função, definida no domínio do tempo e no domínio da frequência. Os componentes de alta frequência são estudados com tempo de resolução mais nítidos do que os componentes de baixa frequência . (DAUBECHIES, 1990).

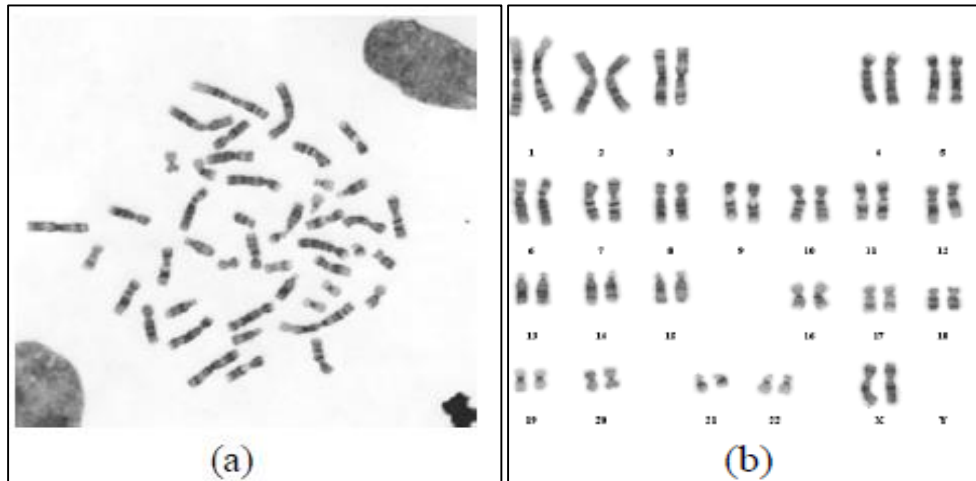


FIGURA 9 – (A) CROMOSSOMO NA METÁFASE, (B) CARIÓTIPO DOS CROMOSSOMOS DA FIGURA 1 (A)

FONTE: XIONG (2007)

Em ABE (2009), é criado um modelo de cromossomo fazendo-se uma segmentação manual, posteriormente o cromossomo é submetido a uma subdivisão, realizada utilizando o algoritmo LBP (*Local Binary Pattern*). O cromossomo é dividido em várias sub-regiões, e o LBP gera um padrão binário para cada região e o conjunto destes padrões binários é um padrão para cada cromossomo. Depois de obtidos os modelos padrões LPB de cada cromossomo, faz-se uma busca na imagem procurando padrões LBP similares aos criados anteriormente. A metodologia é capaz de reconhecer cromossomos que estão sobrepostos e encostados.

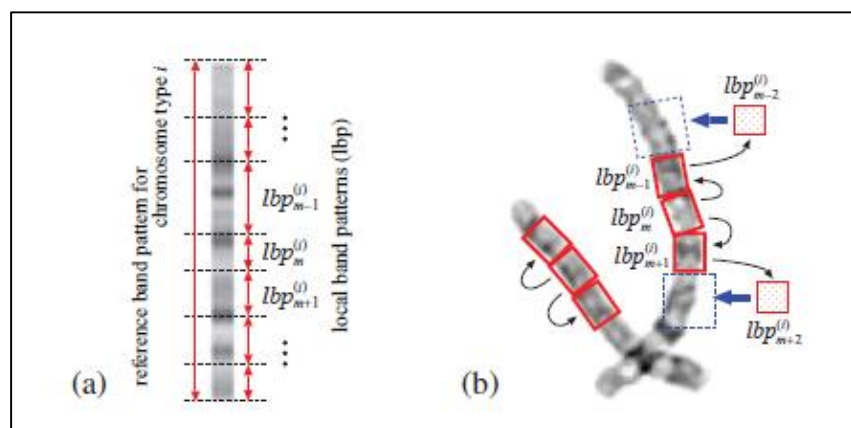


FIGURA 10 – (A) DIVISÃO DO CROMOSSOMO EM SUB-REGIÕES, (B) RECONHECIMENTO DAS SUB-REGIÕES

FONTE: Abe (2009)

A abordagem de KURTZ et al (2008) utilizou a limiarização, dilatação e erosão para criar uma máscara binária, por meio deste processo é possível determinar qual o conjunto de coordenadas a qual o cromossomo pertence e definir os locais a serem segmentados (Figura. 11) a segmentação utilizada foi a de 4-vizinhos que verifica a vizinhança de quatro pixels e armazena os pixels selecionados.

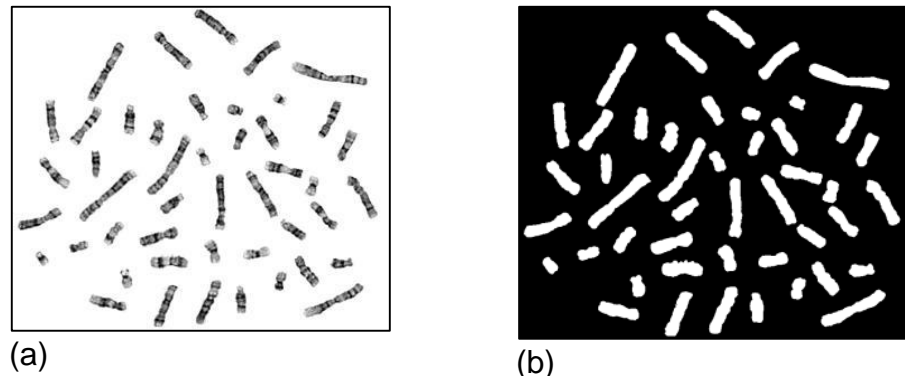


FIGURA 11 – (A) METÁFASE DOS CROMOSSOMOS, (B) MASCARÁ OBTIDA A PARTIR DA LIMIAÇÃO;

FONTE: KURTZ et al (2008)

O reconhecimento dos cromossomos é feito a partir de um modelo concebido chamado Ideograma (Figura 12), é realizado então uma comparação de cada imagem do ideograma com os cromossomos em análise (Figura 12) e o cromossomo que tiver a característica mais similar é identificado, uma característica relevante é a diferença dos níveis de cinza entre a imagem e o ideograma. A figura representa um modelo de cromossomos ordenados por padrões e tamanhos.

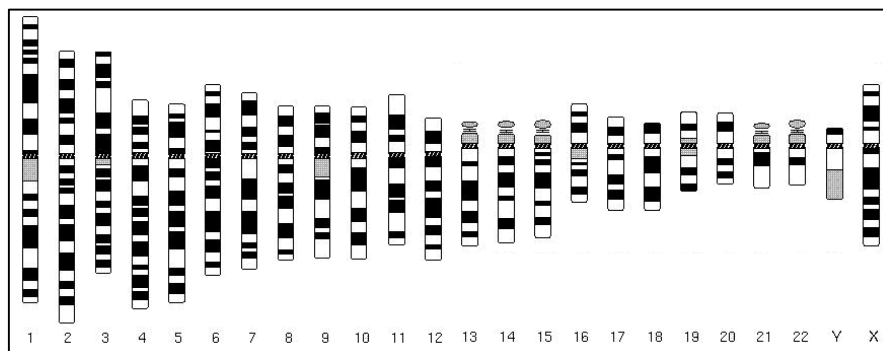


FIGURA 12 – (A) IDEOGRAMA;

FONTE: KURTZ et al (2008)

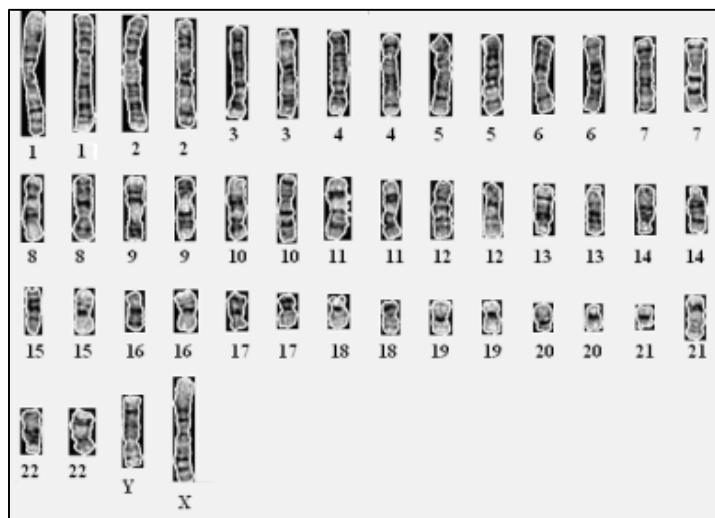


FIGURA 13 – CLASSIFICAÇÃO;

FONTE: KURTZ et al (2008)

A abordagem de GRISAN, POLETTI, RUGGERI (2009) faz a segmentação em imagens de cromossomos que utilizam no padrão bandeamento Q (CASPERSON. 1970). A segmentação da imagem utiliza um *threshold* local adaptativo que divide a imagem em mosaicos de dimensões fixadas e para cada mosaico aplica o limiar de Otsu (OTSU, 1979), com o objetivo de separar melhor a imagem do fundo. Para a remoção de manchas e buracos foi utilizado método de abertura da morfologia matemática. Para a remoção de sobreposições utilizou um método de identificação do eixo principal para verificar se é um único cromossomo ou mais, o algoritmo utilizado é o de vessel-tracking que geralmente é utilizado para verificar bifurcações e cruzamento em imagens de retina. Cromossomos podem apresentar várias formas, configurações e centrômeros nem sempre visíveis, tornando difícil criar um modelo para eles. O método utiliza uma árvore de hipóteses baseada em critérios como curvatura e protuberâncias dos cromossomos (Figura. 15), a borda é percorrida caso houver uma protuberância existe a probabilidade de ser mais que um cromossomo, caso seja mais que um cromossomo através de alguns critérios são gerados candidatos a corte.

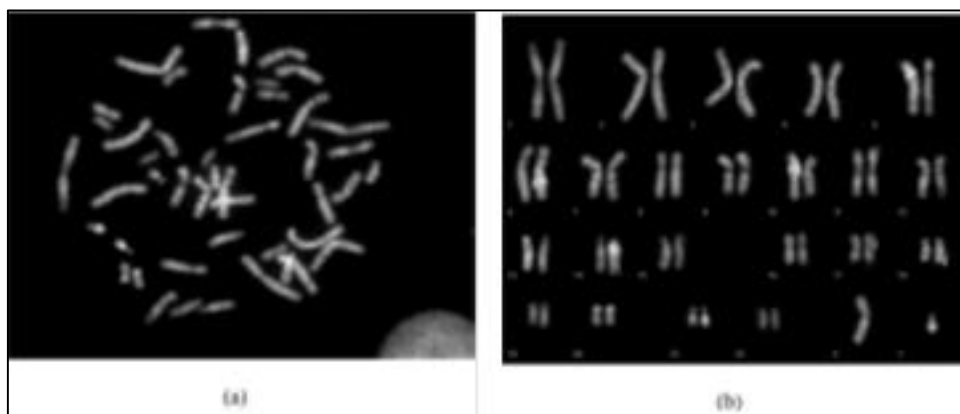


FIGURA 14 – (A) IMAGENS DE CROMOSSOMOS EM BANDEAMENTO Q, (B) CARIÓTIPO

FONTE: GRISAN, POLETTI, RUGGERI (2009)

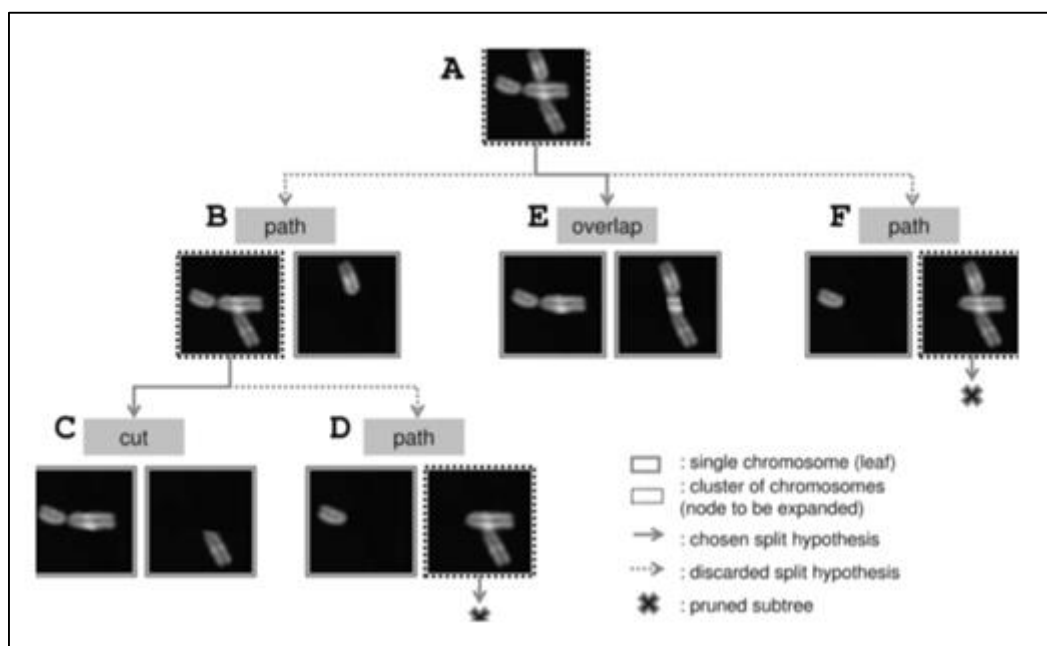


FIGURA 15 – ÁRVORE DE DECISÕES PARA RETIRADA DE SOBREPOSIÇÕES

FONTE: GRISAN, POLETTI, RUGGERI (2009)

Em SOMASUNDARAN, NIRMALA (2010), a metodologia consiste na utilização de uma rede neuronal no processo de comparação de cada cromossomo, foram extraídas características e armazenadas em um banco de dados do sistema biométrico. Os parâmetros são calculados em relação aos pontos que são estáveis. Para cada cromossomo a densidade, valor limiar e posições das bandas são características que são extraídas. Na identificação

cada cromossomo é comparado a vários outros cromossomos, a tomada de decisão vai depender do número máximo de combinações entre a amostra. A maioria dos cromossomos pode ser classificado com base no seu comprimento relativo, índice centromérico, e padrão de bandas; nota-se que o comprimento do cromossomos declina gradualmente na sequência 1-22. Como resultados observou-se que é difícil obter um modelo consistente e generalizável para bandagens de imagens cromossômicas utilizando apenas rede neuronal e amostras biométricas.

Em (FENG et al, 2011), as imagens são pré-processadas e retiradas suas sobreposições e toques, é feita então a esqueletização do cromossomo onde alguns pontos são suprimidos de modo a formar apenas uma linha com o esqueleto do cromossomos. Calcula-se o perfil de densidade dos cromossomos, onde temos o comprimento do eixo mediano, figura 16.

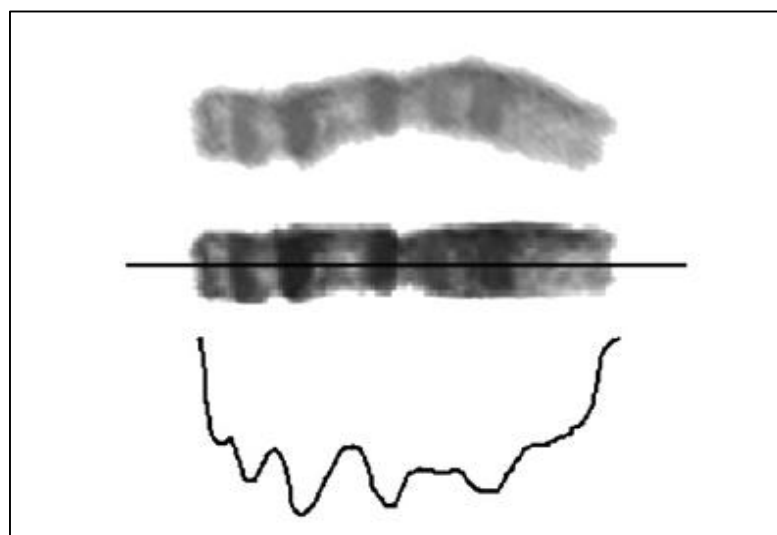


FIGURA 16 – EIXO MÉDIO PRINCIPAL E PERFIL DE DENSIDADE

FONTE: (FENG et al, 2011)

O índice centromérico é medido através da detecção da parte mais fina de um cromossomo. Os ideogramas contêm o padrão de bandas de todos os cromossomos e o seu tamanho proporcional em relação aos outros. A extração das características servem para o reconhecimento dos padrões e posterior identificação dos cromossomos.

Portanto, a partir dos conceitos e definições apresentados no estudo bibliográfico, inicia-se o próximo capítulo que exhibe os procedimentos utilizados para a realização da pesquisa sob uma visão metodológica geral do projeto.

3 MATERIAIS E MÉTODOS

Neste capítulo são apresentados os procedimentos, métodos e metodologia utilizados no desenvolvimento desta pesquisa, além de ambiente de hardware, softwares utilizados e linguagem de programação escolhida para o projeto.

3.1 AMBIENTE HARDWARE

O *hardware* utilizado no desenvolvimento do sistema foi o computador notebook Asus A42F com as seguintes configurações:

- processador: Intel Core I5-460M 2.53 GHZ
- memória: 3 GHz DDR3
- placa gráfica: Intel HD Graphics ;
- disco rígido 500 GB -.

3.2 AMBIENTE SOFTWARE E LINGUAGEM DE PROGRAMAÇÃO

As linguagens de programação utilizadas são Matlab e C, utilizando a biblioteca de processamento de imagens Intel OpenCV (BRADSKY; Gary, KAEHLER, 2008), IBM SPSS, IDE Visual Studio, 2010.

3.3 DESCRIÇÃO DA METODOLOGIA

A metodologia adotada nesta pesquisa é apresentada nas etapas descritas nas próximas subseções, em 3 métodos. Estes métodos foram criados sequencialmente na tentativa de obter os melhores resultados. A metodologia está baseada em Antunes (2003), que afirma que não existe um modelo formal para a segmentação, esse processo é ajustável aos diferentes tipos de imagens, com limiares definidos conforme a complexidade dos alvos investigados.

3.3.1 DEFINIÇÃO DO MODELO GENÉRICO PARA O PROCESSAMENTO DE IMAGENS (PDI)

Os algoritmos desenvolvidos para cada método baseiam-se em similaridades e em descontinuidades, conceitos exibidos no item 2.2. (Técnicas de Processamento de Imagens). As etapas definidas para o modelo consistem na identificação do problema, aquisição da imagem, pré-processamento, segmentação, extração de características, reconhecimento, interpretação e resultado (FACON, 2002) (Figura 17). Estas etapas irão servir de modelo para a aplicação dos três métodos utilizados para a análise das imagens de cromossomos, que serão descritos nos próximos itens.

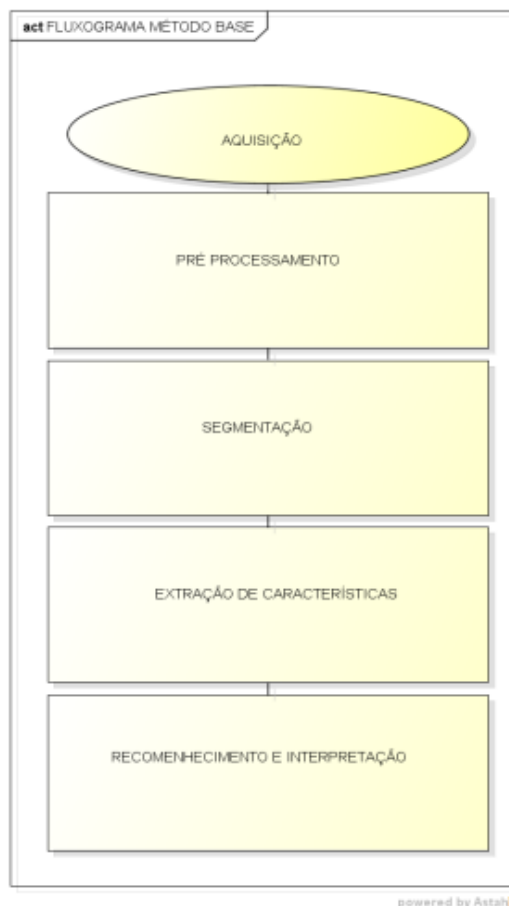


FIGURA 17 – MODELO GENÉRICO PARA O PROCESSAMENTO DE IMAGENS

3.3.2 DEFINIÇÃO DA BASE DE IMAGENS

O presente projeto iniciou utilizando imagens de cariótipos adquiridas do roedor *Akodon montensis* fornecidas em parceria com o banco de imagens em filmes fotográficos, pertencentes ao acervo fotomicrográfico do Laboratório de Citogenética Animal, Departamento de Genética, Setor de Ciências Biológicas da Universidade Federal do Paraná.

A primeira técnica, ainda em fase experimental de definição da metodologia, usava imagens de negativos utilizando um escâner de alta resolução, mas não foi obtido sucesso. Posteriormente, foram escaneadas fotografias, já reveladas, de metáfases de cromossomos da mesma espécie de roedor, em papéis de filmes fotográficos, utilizando técnica de coloração, bandeamento G. As imagens foram adquiridas em níveis de cinza.

Os dois métodos iniciais 1 e 2, descritos na Metodologia (Item 3.5), utilizaram o banco de imagens em filmes fotográficos de cariótipos adquiridas do roedor *Akodon montensis*. No Método 2, além da base de imagem do roedor, foi utilizada a base com cariótipos de humanos (*Homo sapiens*) Zooweb Karyotypes (2013). No Método 3 foi utilizada somente a base com cromossomos humanos (*Homo sapiens*) Zooweb Karyotypes (2013).

3.3.3 MÉTODOS

Foram aplicados três Métodos para análise das imagens dos cromossomos, baseados no Método genérico das etapas de processamento de imagem (PDI) (Figura 17), exibidos no item 3.4.1. As imagens, adquiridas em banco de imagens, são pré-processadas com funções de normalização e aplicação de filtros. Na segmentação, por meio de técnicas de processamento de imagens, somente o conjunto de interesse para extração de características necessárias é separado para o reconhecimento da classe dos cromossomos. Assim, é possível a interpretação do resultado e análise final. A descrição das

particularidades de cada método e seus resultados estão descritos nas subseções a seguir.

3.3.4 MÉTODO 1

Método 1, após a aquisição da imagem e seguindo a estrutura do Método genérico, consiste no aperfeiçoamento da mesma utilizando a técnica de melhoria do contraste, que faz uma equalização de histogramas com reorganização dos pixels para cada nível de cinza (WANG; CHEN; ZHANG, 1999). Foram utilizadas imagens do roedor *Akodon montensis*, pertencentes ao acervo fotomicrográfico do Laboratório de Citogenética Animal. Com maior nitidez da imagem é aplicado um limiar que cria uma máscara binária da imagem. Utilizando a máscara binária como base separa-se o fundo da imagem e os objetos de interesse. Para a rotulação da imagem utiliza-se a técnica de detecção das bordas e o convex hull, para uma criação dos rótulos das imagens.

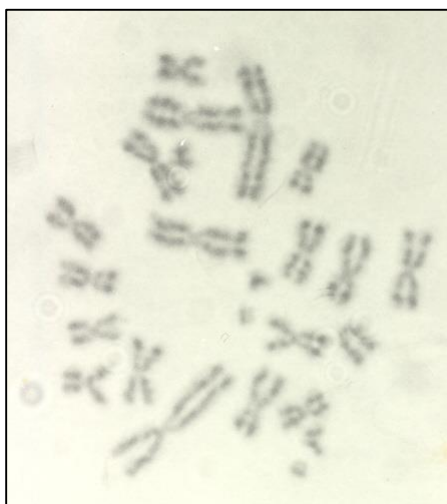
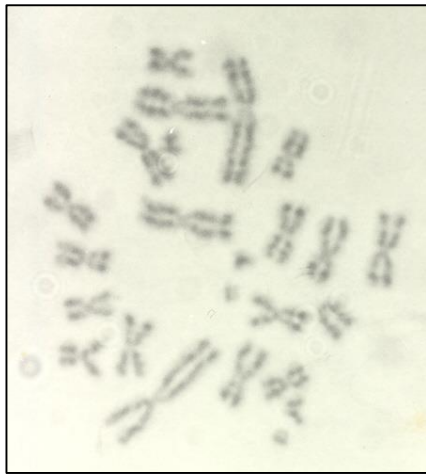


FIGURA 18 – EXEMPLO DAS IMAGENS UTILIZADAS NA METODOLOGIA 1

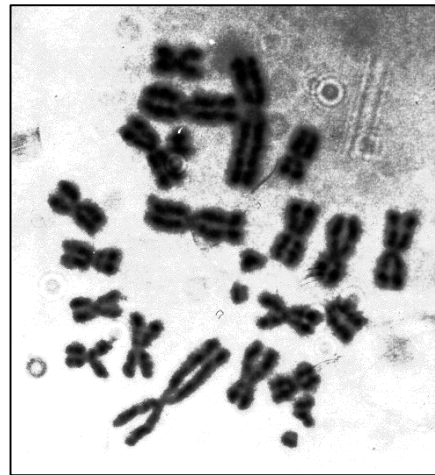
Fonte: Acervo UFPR

A figura 18 acima é a imagem dos cromossomos do roedor *Akodon montensis* em bandeamento G é um exemplo das imagens utilizadas na aplicação do Método 1. As Figuras (a), (b), (c), (d) e (e) exibem o método, utilizando como imagem original a Figura 18.

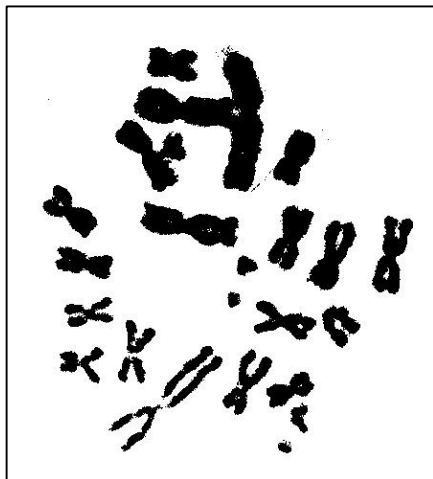
Com o intuito de melhoria da imagem original Figura 19 (a), aplica-se a técnica de Equalização de Sub-imagens (WANG; CHEN; ZHANG, 1999), para melhoria do contraste como mostrado na Figura 19 (b), posteriormente aplica-se um limiar Otsu para uma primeira segmentação da imagem, Figura 19 (c). Utilizando o resultado do limiar como máscara consegue-se separar o objeto de interesse do fundo da imagem. Como tentativa de extração de informações da imagem aplica-se a operação de detecção do contorno e posteriormente a aplicação do algoritmo de ponto convexo a fim de criar rótulos.



(a) Imagem Original



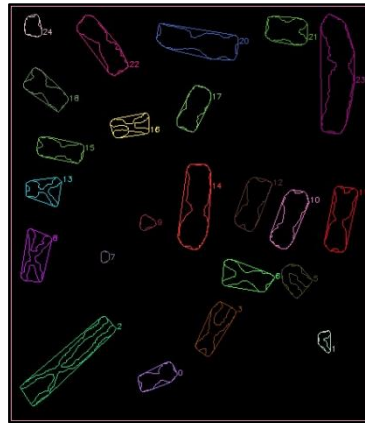
(b) Melhoria do contraste



(c) Limiarização da imagem



(d) Segmentação do objeto de interesse



(e) Criação de rótulos
FIGURA 19 – FLUXO DO MÉTODO 1

Passo 1. Equalização de Sub-Imagens

A fim de trazer uma melhoria no processamento, uma etapa importante é melhoria do contraste da imagem, a qual consiste na reorganização dos níveis de cinza da imagem.

Passo 2. Limiar de Ostu

O algoritmo de Ostu, faz a binarização da imagens para que haja a identificação dos objetos na imagem.

Passo 3. Detecção de Bordas

Na detecção de bordas, ocorre uma mudança abrupta dos níveis de cinza da imagem e do fundo da imagem. Assim o algoritmo localiza as bordas na imagem.

Passo 4. Convex Hull

Neste caso são as linhas que contornam as bordas na imagem mostrado na imagem abaixo. Utilizou-se para reconhecimento de objetos em uma imagem para posterior segmentação.

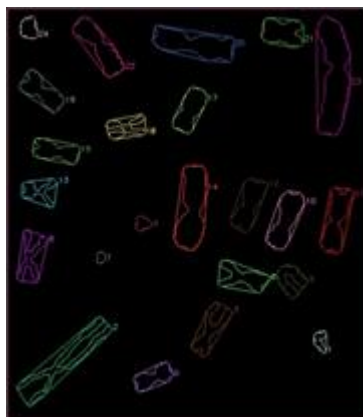


FIGURA 20 – RÓTULOS CRIADOS

Este processamento deve resultar na determinação das características pela segmentação correta dos objetos. Para que o algoritmo convex hull funcione de forma eficaz, não deve haver ruído nas imagens.

3.3.5 MÉTODO 2

O Método 2 consiste na adaptação do Modelo genérico na fase da segmentação. Com base em formas pré-definidas na criação de modelos de forma que possam ser mais robustos, o método originalmente foi criado para a segmentação do dente em imagens de raio-X (LIRA; GIRALDI; NEVES, 2010). Foram utilizadas dois tipos de imagens, as pertencentes ao acervo fotomicrográfico do Laboratório de Citogenética Animal, com cromossomos de roedores; e as Imagens do Zooweb Karyotypes (2013), com cromossomos de humanos. A inserção da nova base de imagens da internet foi justificada, pois as imagens anteriores tinham complexidade elevada, o que poderia prejudicar o desempenho do algoritmo proposto.

O Método 2 baseou-se em um algoritmo desenvolvido em Matlab para segmentação e extração do dente utilizando como base, imagens de radiografias. No primeiro passo do algoritmo do Método 2 faz-se a extração dos contornos individuais, que utiliza o método 1 e aplica-se a segmentação manual. No segundo passo do algoritmo cria-se um modelo baseado em formas do cromossomo, submete-se 30 imagens dos contornos do cromossomo 1 para

parametrização com o PCA (Análise da componente principal), então cria-se um modelo para o cromossomo 1 (Shape Model), Figura 21. Utiliza-se este modelo para procurar os cromossomos similares na imagem original. O modelo de treino é submetido ao reconhecimento na imagem.



FIGURA 21 – MODELO DO CROMOSSOMO 1

Este processamento deve resultar na determinação das características por similaridade com os cromossomos no conjunto de reconhecimento, que será ordenado de forma numérica.

3.3.6 MÉTODO 3

O Método 3 utilizou-se somente de imagens da base de dados de humanos Zooweb Karyotypes (2013), disponibilizadas em grupos e são de cariótipos em metáfases cromossômicas. Na primeira etapa, para obter-se mais informações após a aquisição da imagens, aplica-se escala de 200%. Na sequência faz-se a limiarização utilizando o algoritmo LAT “Local Average Thresholding”. Com objetivo de preencher os orifícios do processo de limiarização utilizou-se a morfologia matemática, a operação abertura binária, com o elemento estruturante quadrado. Fez-se a inversão de imagem para aplicação do algoritmo de Rotulação Binária, onde cada cromossomo é segmentado da imagem para análise e reconhecimento. As características necessárias para o reconhecimento da classe de todos os cromossomos foram cálculo da área e perímetro, cálculo da altura através do afinamento de cada cromossomo, fator de compacidade, estas características foram utilizadas para

os resultados de 3/1. Para a gerar as reconhecimento do método 3/2 foi utilizado o histograma LBP (*Local Binary Pattern*) das sub-imagens como característica. A visão geral do método 3 está exibida no fluxograma, figura 22.

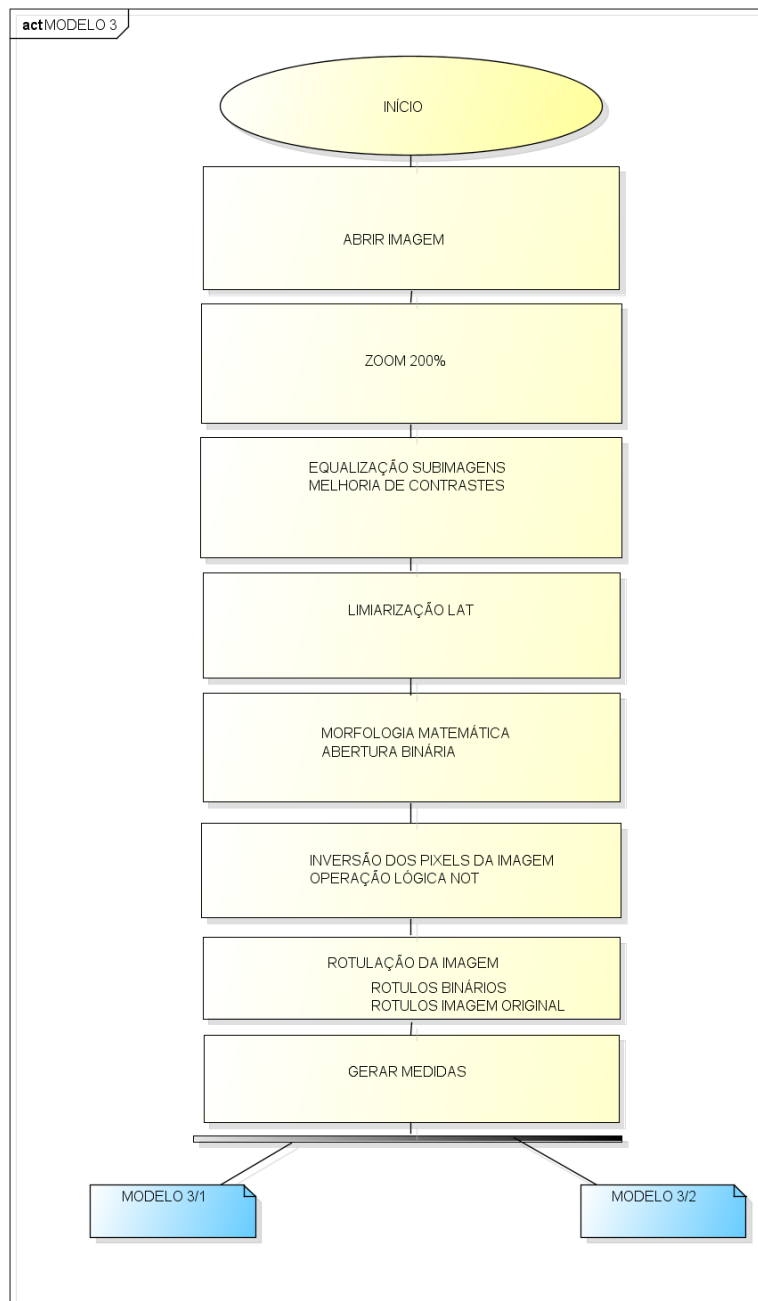


FIGURA 22 – VISÃO GERAL DO MODELO 3

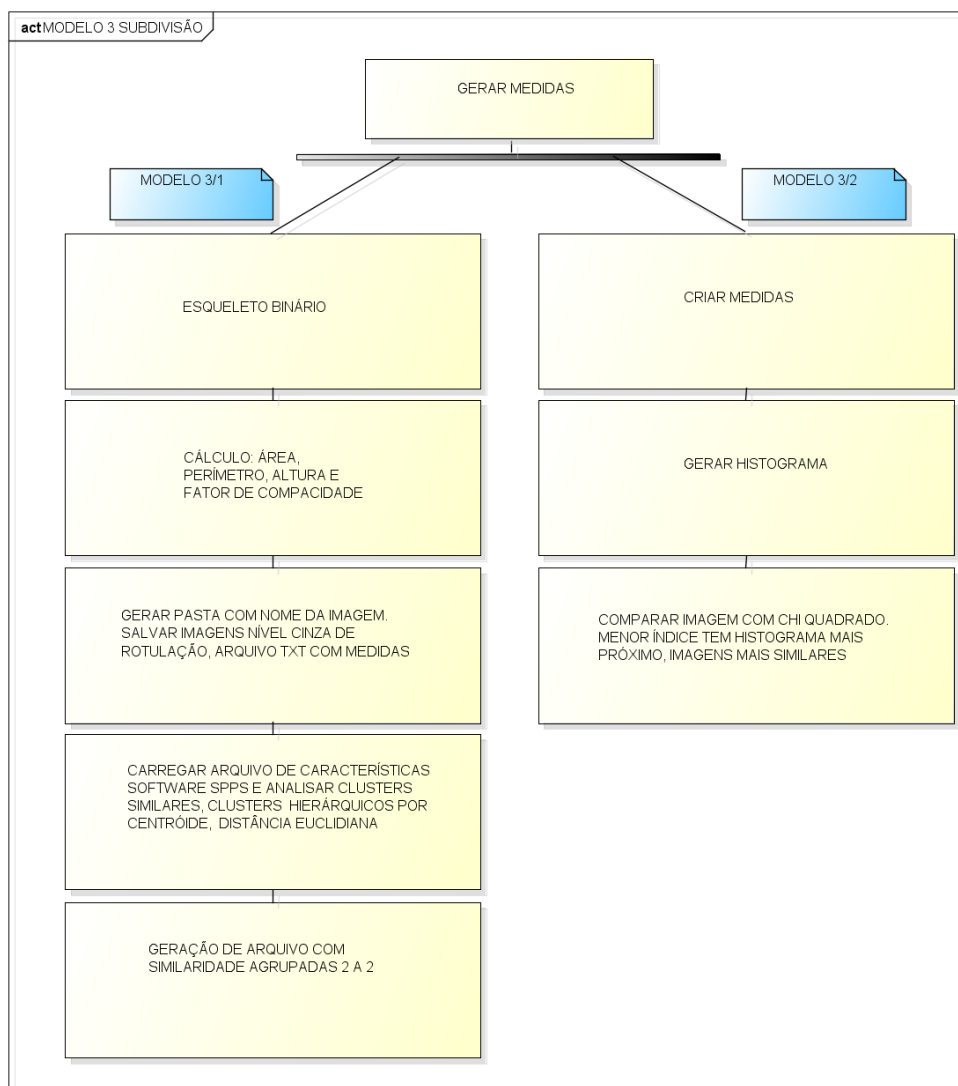


FIGURA 23 – VISÃO GERAL DO MODELO 3 - SUBDIVISÃO

Os agrupamentos dos cromossomos são separados da seguinte forma:

- Grupo A: pares 1, 2 e 3
- Grupo B: pares 4 e 5
- Grupo C: pares 6 ao 12, mais o cromossomo X
- Grupo D: pares 13, 14 e 15
- Grupo E: pares 16, 17 e 18
- Grupo F: pares 19 e 20
- Grupo G: pares 21, 22 e Y

Estes grupos são formados de acordo com as características de área, perímetro e altura. Mais detalhes de cada etapa do Método 3 são apresentados a seguir.

Nesta etapa são apresentados os passos utilizados no desenvolvimento, Passos 1. ESCALA; 2. Equalização de Sub-imagens; 3. Limiarização de LAT (Local Average Thresholding); 4. Morfologia matemática - Abertura binária; 5. Inversão da Imagem binária; 6. Rotulação binária; 7. Segmentação dos Rótulos Binários; Segmentação da Imagem Original e 8. Gerar Medidas.

Passo 1. ESCALA

Neste método utiliza-se uma escala na imagem em 200%, que é suficiente para executar melhor a quantidade de informações presentes na imagem

Passo 2. Equalização de Sub-imagens

O objetivo do algoritmo é melhoria do contraste da imagem.

Passo 3. Limiarização de LAT (Local Average Thresholding)

O algoritmo de LAT, faz a binarização da imagens para que haja a identificação dos objetos na imagem.

Passo 4. Morfologia matemática - Abertura binária

Este processo é utilizado projeto a fim de fechar os “buracos” deixados pela limiarização, assim evita-se perder alguma informação da imagem.

Passo 5. Inversão da Imagem binária

Inversão da imagem binária ou operação de NOT é o processo onde há inversão dos valores. A imagem base contém somente pixels pretos e brancos, os pixels pretos são convertidos em brancos e os brancos são convertidos em pretos. Este processo é importante para que o algoritmo de Rotulação Binária funcione corretamente.

Passo 6. Rotulação binária

O próximo passo é a rotulação dos pixels em uma imagem binária utilizando um rótulo R.

Para a visualização o algoritmo de rotulação identifica cada rótulo encontrado e o representa através de uma cor aleatória, mas internamente no software desenvolvido são criadas múltiplas sub-imagens.



FIGURA 24 – IDENTIFICAÇÃO DO RÓTULO BINÁRIO

Através das informações adquiridas na rotulação recupera-se a imagem original e criam-se rótulos da imagem original em níveis de cinza.

Passo 6.1. Segmentação dos Rótulos Binários

O algoritmo de rotulação binária gera rótulos de coordenadas X e Y de cada rótulo, a partir das coordenadas geradas pelo algoritmo de rotulação, são construídas as sub-imagens. Este processo é feito para que se tenham as sub-imagens e não suas coordenadas, o objetivo é facilitar os futuros processamentos de imagens.

Passo 6.2. Segmentação da Imagem Original

Assim como a segmentação dos rótulos binários, o qual possui informações das coordenadas X e Y dos rótulos, constrói-se a sub-imagem cinza a partir da imagem original cinza. Este processo é realizado para todos os rótulos. O software desenvolvido, salva as sub-imagens em disco para posterior análise.

Passo 7. Gerar Medidas

Os cálculos das medidas de área, perímetro e fator compacidade, deram-se sobre a imagem binária, a medida da altura foi calculada verificando-se a altura do esqueleto. Para o cálculo da área número de pixels das sub-imagens são computados, para o cálculo do perímetro é medido a dimensão do contorno de cada sub-imagem e o cálculo do fator compacidade, o qual faz o cálculo de quanto a imagem é alongada ou arredondada. Onde P^2 é o perímetro e S a área.

$$Fc = \frac{P^2}{4 * \pi * S} \quad (9)$$

FONTE: FELIPE , NEVES (2008)

O objetivo destes cálculos é discriminar o cromossomo através das medidas.

A partir deste ponto, na Segmentação, Modelo 3 foi subdividido em mais duas etapas, que vão gerar resultados diferentes:

3.3.6.1 MODELO 3/1

O Modelo 3/1 vai se diferenciar do Modelo 3/2 na Extração de Características, conforme exibido na figura 25.

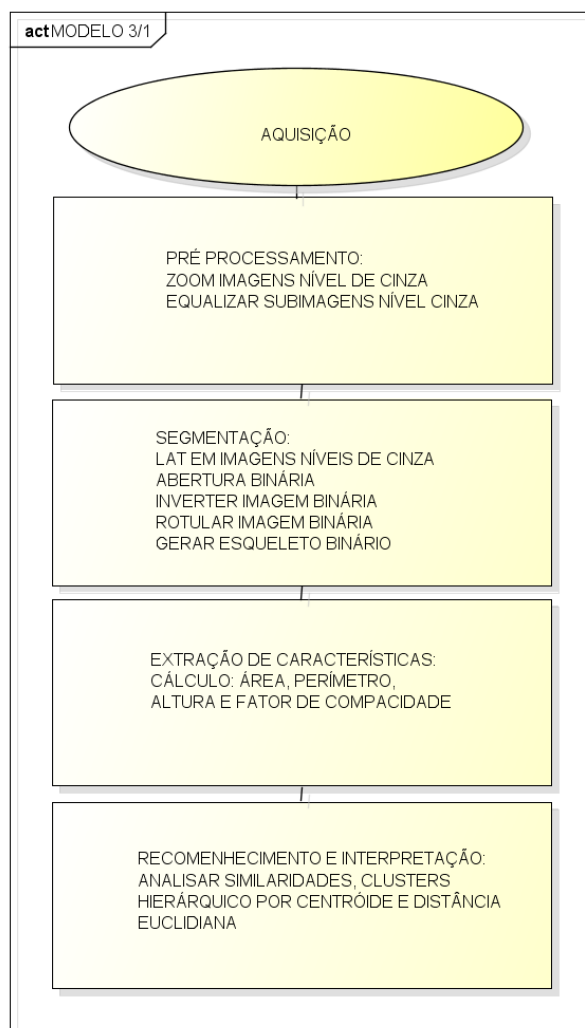


FIGURA 25 – VISÃO GERAL SUBDIVISÃO MODELO 3/1

Passo 1. Esqueletização Binária

A esqueletização consiste em uma operação morfológica onde se obtém apenas uma representação compacta da imagem, geralmente em uma linha contínua.

O algoritmo itera e caso não haja mais pixels a serem removidos a esqueletização está completa. Imagem binária 22, imagem do esqueleto 23.

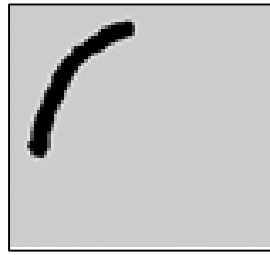


FIGURA 26 – IMAGEM BINÁRIA

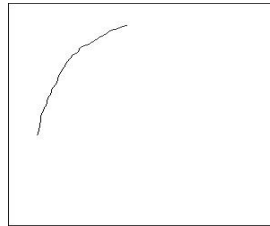


FIGURA 27 – IMAGEM DO ESQUELETO

Passo 2. Cálculo de Medidas Relativas

As medidas extraídas são medidas relativas, pois alternando de imagem para imagem tinha-se medidas diferentes. A medida relativa significa o quanto por cento as medidas de um objeto tem significância em relação aos outros objetos.

- Área;
- Perímetro;
- Altura;
- Fator de Compacidade

Geração de pasta e Arquivamento

Nesta etapa gera-se pasta com nome da imagem. Salva-se as Sub-imagens em nível de cinza e arquivos .txt com todas as medidas.

Passo 3. Arquivamento de Características, Software SPSS e Análise Clusters

A análise utiliza-se do software IBM – SPSS, na função de análise hierárquica de clusters, clusterização por centroide e distância euclidiana ao quadrado.

Carrega-se o Arquivo de Características, utiliza-se o software SPSS e analisa-se as similaridades dos clusters.

Arquivo com Similaridade gerado pelo SPSS

O software SPSS Gera uma análise de similaridades agrupadas 2 a 2.

3.3.6.2 MODELO 3/2

O Modelo 3/2 tem sua metodologia modificada do Modelo 3/1 na Extração de Características, conforme exibido na figura 28. Neste Modelo a imagem LBP faz parte do processo de Segmentação, a Extração das Características é realizada pelo Histograma LBP e no Reconhecimento e interpretação é utilizado o CHI quadrado.

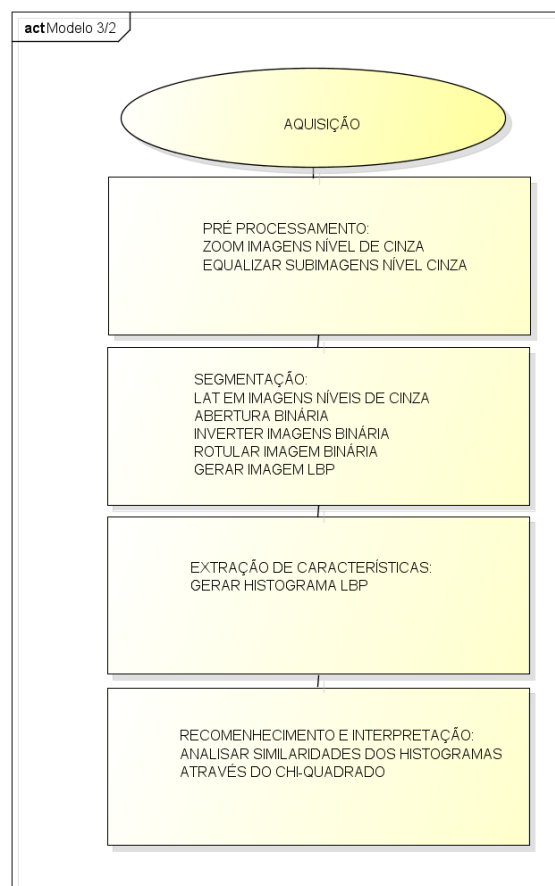


FIGURA 28 – VISÃO GERAL SUBDIVISÃO MODELO 3/2

Passo 1. Imagem LBP

A imagem em tons de cinza abaixo foi submetida ao algoritmo. Figura. 29 imagem original, Figura. 30 imagem LBP.

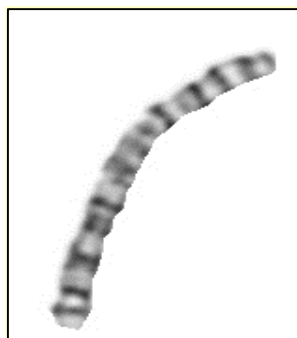


FIGURA 29 – IMAGEM SEGMENTADA



FIGURA 30 – IMAGEM LBP

Passo 2. Histograma

Para cada imagem no padrão LBP gera-se um Histograma. Estes histogramas têm as informações das distribuições de frequência dos pixels das imagens.

Passo 3. Comparar Histogramas (Chi-quadrado)

Para comparação de histogramas utiliza-se o coeficiente chi-quadrado, que analisa um valor de dispersão entre duas variáveis.

Os resultados obtidos com a aplicação dos três Métodos descritos são apresentados no capítulo a seguir.

4 RESULTADOS

Os resultados obtidos com o projeto serão exibidos neste capítulo. O propósito do projeto é fazer uma análise comparativa dos métodos, a ser realizada no capítulo Discussão.

4.1 MÉTODO 1

A complexidade no processamento de imagens foi alta, pois não se conseguiu determinar características para segmentação correta dos objetos. Houve perda de informações importantes nas imagens, não sendo possível uma rotulação eficaz, por este motivo a reconstrução da imagem do cromossomo foi inviável, como mostra a imagem Figura 31. Observando a figura 31 verificamos que devido à baixa qualidade das imagens, o algoritmo não obteve sucesso e criou inúmeros rótulos inclusive com ruídos. Os mesmo testes foram realizados com o Banco de Imagens de humanos, os resultados também não foram eficazes. O motivo principal desta ineficácia é pelo algoritmo de limiarização Otsu não ser o ideal para este novo banco de imagens. Nas execuções do software sobre os dois bancos de imagens, primeiro são reconhecidas as bordas dos cromossomos e posteriormente a rotulação pelo convex hull, a qual cria um contorno ao redor de cada imagem. Devido à baixa qualidade da imagem perdem-se referências importantes e não se consegue executar o algoritmo convex hull de forma eficaz. Outro problema é que temos toques entre os cromossomos, sobreposições, problemas estes que necessitariam de um estudo mais aprofundado para tratamento específico destes casos.

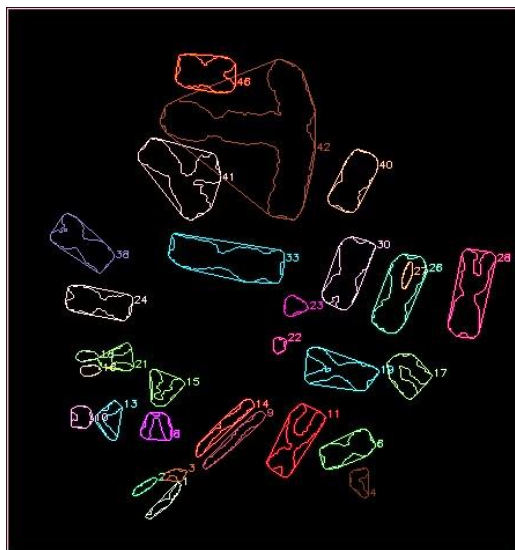


FIGURA 31 – CRIAÇÃO DE RÓTULOS

4.2 MÉTODO 2

Esse processo não foi eficaz, pois o modelo encontrou pouca similaridade com os cromossomos no conjunto de reconhecimento, as similaridades com o cromossomo 1 foram ordenadas de forma numérica. Apesar do algoritmo proposto ter sido modificado para receber o novo conjunto de imagens, seriam necessárias novas funcionalidades e mais refinamento na análise de seus métodos para saber se seus resultados seriam satisfatórios. Por este motivo, não foi dada continuidade a sua implementação.

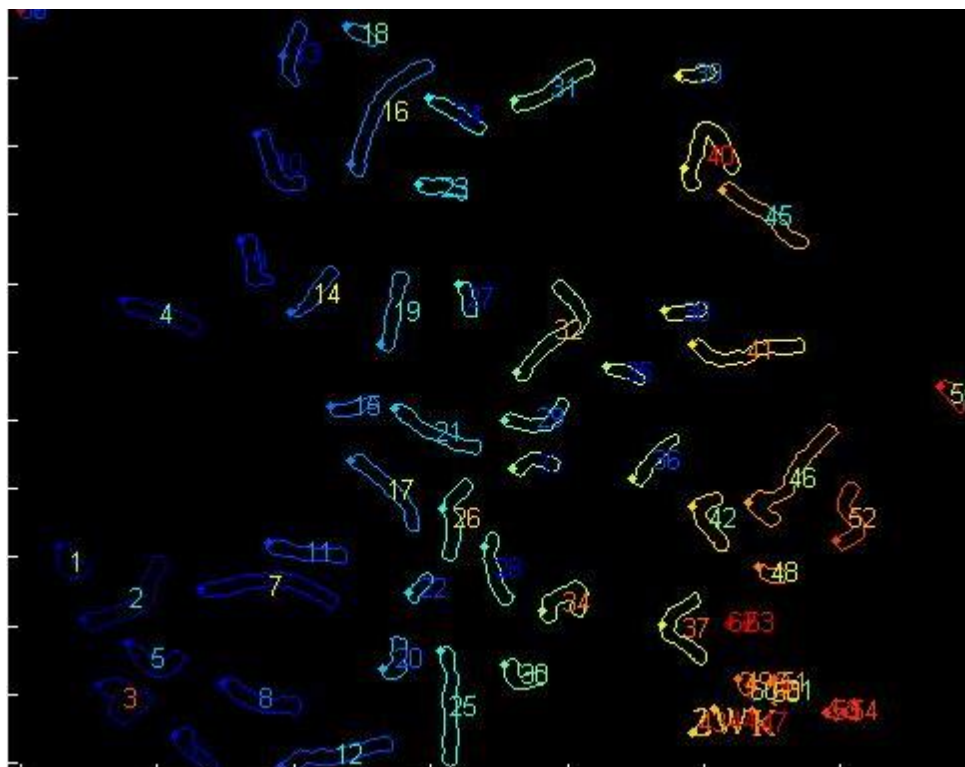


FIGURA 32 – RECONHECIMENTO

4.3 MÉTODO 3

O método 3 conforme descrito anteriormente é baseado no primeiro método, utilizando um novo banco de imagens, com 30 imagens. O comportamento desta base está descrita a seguir.

A metodologia está sendo construída em 10 etapas que foram descritas anteriormente. A seguir são apresentadas as imagens obtidas durante o processo:

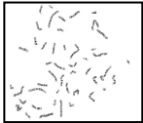
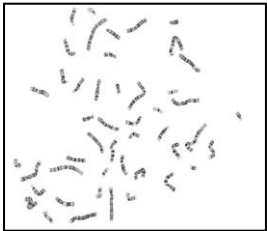

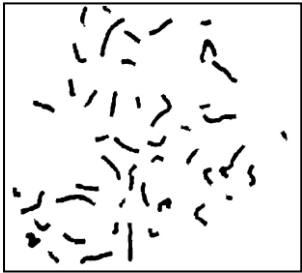
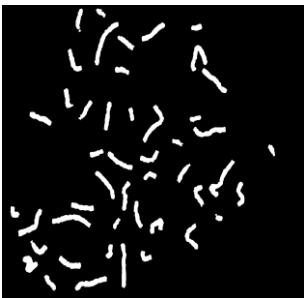
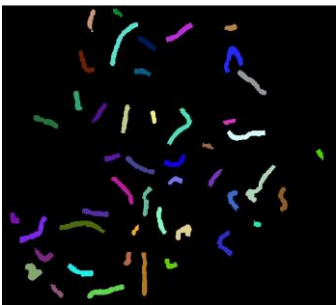
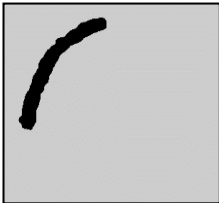
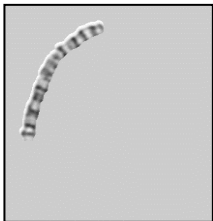
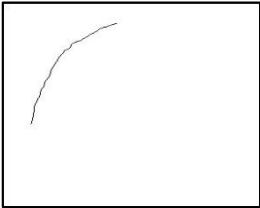
 <p>(a) Imagem Original</p>	 <p>Imagem com ESCALA</p>
 <p>(b) Imagem Limiarizada com LAT</p>	 <p>(c) Imagem com abertura morfológica binária</p>
 <p>(d) NOT da imagem</p>	 <p>(e) Rotulação binária</p>
 <p>(f) Rotulo binário segmentado</p>	 <p>(g) Imagem segmentada</p>
 <p>(h) Esqueleto binário da imagem (g)</p>	

FIGURA 33 – ETAPAS DE PROCESSO

A Figura 33 com as etapas dos processos, é explicada como a imagem original (a) consiste em um exemplo do conjunto de imagens analisadas, adquiridas do banco Zooweb Karyotypes (2013); a imagem com Escala (b) é o processo de ampliação de (a) para a melhoria de riqueza em informações para posteriores processos; a imagem Limiarizada LAT (c), é um processo de segmentação onde consegue-se separar o fundo da imagem, o objetivo da imagem binarizada é conseguir delimitar a quantidade de informações na imagem, neste caso o objeto em estudo fica sendo o pixel preto e o fundo branco; a abertura morfológica binária (d), é o processo onde consegue-se reconstruir algumas informações que foram perdidas pelo processo de limiarização, como o exemplo do processo de (c) para (d) onde os orifícios foram preenchidos através de operação de dilatação e erosão; a inversão da imagem (e) é o processo onde faz-se uma inversão de pixel ou uma operação de NOT na imagem, ou seja pixel brancos tornam-se pretos e pretos tornam-se brancos; a Rotulação binária (f) é o processo onde cada um dos objetos da imagem são separados, então cria-se uma imagem para cada objeto; o Rotulo binário segmentado (g), representa um dos casos de rotulação; o Rotulo cinza segmentado (h), representa um dos casos de rotulação; o Esqueleto binário (i), consiste no processo de afinar a imagem (g) e representá-la em forma de somente a linha.

4.3.1 CÁLCULO DAS CARACTERÍSTICAS

O vetor de característica de cada cromossomo é apresentada a seguir. Cada linha contém as medidas de Área, Perímetro, Altura, Fator Compacidade de um rótulo representado como sequencial 1,2,...n. As características geradas são representadas na tabela 6:

TABELA 6 – CARACTERÍSTICAS GERADAS

Área	Perímetro	Altura	Compacidade	Sequencial
0,851842	1,12557	1,01813	2,8299	1
1,62772	1,86868	2,01142	1,96187	2

Área	Perímetro	Altura	Compacidade	Sequencial
4,05139	3,80872	4,27117	1,17545	3
2,61047	2,48555	2,48324	1,77842	4
1,0439	1,16251	0,769804	3,25107	5
1,73617	1,82666	1,88726	2,18995	6
3,555	3,2925	3,79935	1,38022	7
2,27031	2,18427	2,18525	2,00277	8
1,38932	1,50963	1,29128	2,5658	9
3,01943	2,82646	2,83089	1,59074	10
1,76866	1,77489	1,86243	2,36298	11
1,62963	1,84163	1,73827	2,02228	12
2,18765	2,16818	2,33424	1,95861	13
3,65915	3,57214	4,04768	1,20693	14
0,948827	1,1329	0,744971	3,11148	15
2,43465	2,22615	2,18525	2,0677	16
0,942139	1,17366	0,893966	2,87868	17
3,27264	3,19917	3,6007	1,34581	18
0,850409	1,18018	1,06779	2,56973	19
0,863786	1,0552	0,819469	3,2651	20
1,26893	1,43351	1,14229	2,59891	21
1,93062	2,09813	2,18525	1,84582	22
2,56795	2,62168	2,6819	1,57249	23
4,01412	3,69521	4,04768	1,23729	24
1,60957	1,77588	1,71343	2,14804	25
2,67688	2,6658	2,6074	1,58538	26
1,25698	1,4469	0,918798	2,52703	27
2,33385	2,37105	2,33424	1,74723	28
2,17858	2,16534	2,06109	1,9556	29
2,08685	2,14846	2,18525	1,90282	30
2,08398	2,12122	2,21008	1,94932	31
2,21871	2,22549	2,30941	1,88541	32
1,24743	1,28591	0,769804	3,17507	33
3,39781	3,09378	3,30271	1,4941	34
0,67555	0,974144	0,62081	2,9962	35
3,97399	3,72813	4,54433	1,20339	36
0,771579	1,04788	0,893966	2,95746	37
2,29849	2,07472	1,93693	2,24742	38
2,55266	2,62662	2,75639	1,55725	39
2,04767	2,03923	1,88726	2,07246	40
3,32902	3,13817	3,70002	1,42273	41
1,22593	1,32047	0,94363	2,95916	42
1,35827	1,48272	1,24162	2,60033	43
2,87849	2,02585	2,01142	2,95197	44
2,48195	2,29382	2,18525	1,98534	45

Área	Perímetro	Altura	Compacidade	Sequencial
1,44331	1,58066	1,46511	2,43133	46
3,37775	3,1046	3,50137	1,47495	47

Cada linha da tabela representa medidas geradas pelo software desenvolvido, as medidas são representadas proporcionalmente em percentual. Podemos observar na tabela 6, que as linhas 2 e 12 demarcadas tem grande similaridade de medidas, esta similaridade foi apontada pela análise que será explicada neste capítulo.

A Tabela 7 gerada através do software SPSS a qual mostra os agrupamentos dos clusters. Utiliza-se neste software a opção de classificação cluster hierárquico e configuração de aproximação por centroide (centro de massa) e pelo quadrado da distância euclidiana. A tabela abaixo representa quais características da tabela 7 (anterior) obtiveram as melhores aproximações. As aproximações são apresentadas de acordo com a classificação e coeficiente de aproximação, quanto mais próxima a vizinhança menor o índice.

TABELA 7 – CLASSIFICAÇÃO

Área			
Estágio	Combinação do Cluster		Coeficientes
	Cluster 1	Cluster 2	
1	1	19	,000
2	2	12	,000
3	30	31	,000
4	15	17	,000
5	13	29	,000
6	27	33	,000
7	1	20	,000
8	23	39	,000
9	21	27	,000
10	34	47	,000
11	2	25	,000
12	8	38	,000
13	9	43	,000
14	6	11	,000
15	3	24	,000
16	21	42	,000
17	13	32	,000

Estágio	Área		
	Combinação do Cluster		Coeficientes
	Cluster 1	Cluster 2	
18	30	40	,000
19	16	45	,001
20	18	41	,001
21	8	28	,001
22	4	23	,001
23	3	36	,001
24	9	46	,002
25	35	37	,002
26	7	14	,002
27	5	15	,003
28	4	26	,004
29	18	34	,004
30	10	44	,005
31	22	30	,006
32	8	13	,008
33	2	6	,010
34	1	35	,012
35	4	16	,014
36	9	21	,017
37	1	5	,022
38	7	18	,030
39	8	22	,040
40	4	10	,060
41	2	9	,094
42	3	7	,153
43	4	8	,246
44	1	2	,395
45	3	4	1,211
46	1	3	3,697

O Agrupamento cromossômico A é representado por esses três pares, numerados de 1 a 6, os índices acima 36, 24, 14, 3, 7 e 18 são referentes a o número do rótulo gerado e da linha de características. Através dos dados obtidos pôde-se identificar que os seis cromossomos figura 34 foram todos classificados, reconhecidos.




36 24	14 3	7 18
		
1 2	3 4	5 6
Par 1	Par 2	Par 3

FIGURA 34 – CROMOSSOMOS CLASSIFICADOS

Para validação dos resultados foi construída uma tabela do Microsoft Excel, que são representados da seguinte forma:

- Linha é representada sequencialmente pelo SPSS;
- C1 e C2 são os clusteres agrupados, onde seus valores são as linhas de características;
- Coeficiente é o coeficiente de aproximação gerado pela vizinhança;
- Grupo C1 significa a qual grupo pertence o cromossomo que possui a característica c1;
- Grupo C2 significa a qual grupo pertence o cromossomo que possui a característica c2;
- Verificação, é 1 caso seja verdadeiro e 0 caso seja falso;
- Acerto significa qual grupo foi acertado.

Como pode-se observar, todos os cromossomos do Grupo A foram encontrados. Na tabela 8 foi somente analisada a área, a qual obteve melhor resultado.

TABELA 8 – RESULTADOS

SOMENTE ÁREA							
Linha	C1	C2	COEFICIENTE	GRUPO C1	GRUPO C2	VERIFICAÇÃO	ACERTO
15	3	24	0,000	A	A	1	A
23	3	36	0,001	A	A	1	A
26	7	14	0,002	A	A	1	A
38	7	18	0,030	A	A	1	A
42	3	7	0,153	A	A	1	A

A tabela a seguir exhibe a validação de todos os resultados, para este conjunto de dados foi obtido 76% de acerto.

TABELA 9 – VALIDAÇÃO DE TODOS OS RESULTADOS

	C1	C2	COEFICIENTE	GRUPO C1	GRUPO C2	VERIFICAÇÃO	ACERTO
1	1	19	0,000	G	F	0	0
2	2	12	0,000	D	D	1	D
3	30	31	0,000	C	C	1	C
4	15	17	0,000	F	F	1	F
5	13	29	0,000	C	C	1	C
6	27	33	0,000	E	F	0	0
7	1	20	0,000	G	G	1	G
8	23	39	0,000	C	C	1	C
9	21	27	0,000	E	E	1	E
10	34	47	0,000	B	B	1	B
11	2	25	0,000	D	D	1	D
12	8	38	0,000	C	C	1	C
13	9	43	0,000	E	E	1	E
14	6	11	0,000	D	D	1	D
15	3	24	0,000	A	A	1	A
16	21	42	0,000	E	E	1	E
17	13	32	0,000	C	C	1	C

	C1	C2	COEFICIENTE	GRUPO C1	GRUPO C2	VERIFICAÇÃO	ACERTO
18	30	40	0,000	C	D	0	0
19	16	45	0,001	C	C	1	C
20	18	41	0,001	A	B	0	0
21	8	28	0,001	C	C	1	C
22	4	23	0,001	C	C	1	C
23	3	36	0,001	A	A	1	A
24	9	46	0,002	E	E	1	E
25	35	37	0,002	G	G	1	G
26	7	14	0,002	A	A	1	A
27	5	15	0,003	G	F	0	0
28	4	26	0,004	C	B	0	0
29	18	34	0,004	A	B	0	0
30	10	44	0,005	C	C	1	C
31	22	30	0,006	C	C	1	C
32	8	13	0,008	C	C	1	C
33	2	6	0,010	D	D	1	D
34	1	35	0,012	G	G	1	G
35	4	16	0,014	C	C	1	C
36	9	21	0,017	E	E	1	E
37	1	5	0,022	G	G	1	G
38	7	18	0,030	A	A	1	A
39	8	22	0,040	C	C	1	C
40	4	10	0,060	C	C	1	C
41	2	9	0,094	D	E	0	0
42	3	7	0,153	A	A	1	A
43	4	8	0,246	C	C	1	C
44	1	2	0,395	G	D	0	0
45	3	4	1,211	A	C	0	0
46	1	3	3,697	G	A	0	0

A tabela abaixo mostra os resultados da análise com todas as medidas, Área, perímetro, altura e fator compacidade.

TABELA 10 – VALIDAÇÃO COM TODAS AS MEDIDAS

TODOS							
	C1	C2	COEFICIENTE	GRUPO C1	GRUPO C2	VERIFICAÇÃO	ACERTO
1	23	39	0,000	C	C	1	C
2	9	43	0,001	E	E	1	E
3	30	31	0,001	C	C	1	C
4	9	21	0,002	E	E	1	E
5	23	26	0,002	C	B	0	0
6	13	32	0,002	C	C	1	C
7	13	30	0,002	C	C	1	C
8	8	29	0,002	C	C	1	C
9	16	45	0,002	C	C	1	C
10	8	13	0,002	C	C	1	C
11	1	17	0,003	G	F	0	0
12	34	47	0,003	B	B	1	B
13	18	41	0,003	A	B	0	0
14	8	22	0,003	C	C	1	C
15	18	34	0,004	A	B	0	0
16	8	16	0,004	C	C	1	C
17	6	25	0,004	D	D	1	D
18	6	12	0,004	D	D	1	D
19	9	27	0,004	E	E	1	E
20	5	20	0,004	G	G	1	G
21	5	15	0,005	G	F	0	0
22	2	6	0,006	D	D	1	D
23	3	24	0,006	A	A	1	A
24	1	37	0,006	G	G	1	G
25	3	36	0,006	A	A	1	A
26	1	35	0,007	G	G	1	G
27	5	33	0,007	G	F	0	0
28	9	46	0,007	E	E	1	E
29	2	11	0,007	D	D	1	D
30	8	28	0,008	C	C	1	C

TODOS							
	C1	C2	COEFICIENTE	GRUPO C1	GRUPO C2	VERIFICAÇÃO	ACERTO
31	7	18	0,009	A	A	1	A
32	8	40	0,009	C	D	0	0
33	4	8	0,010	C	C	1	C
34	1	5	0,011	G	G	1	G
35	1	42	0,011	G	E	0	0
36	4	38	0,013	C	C	1	C
37	3	14	0,013	A	A	1	A
38	4	23	0,014	C	C	1	C
39	1	19	0,016	G	F	0	0
40	4	10	0,017	C	C	1	C
41	2	4	0,017	D	C	0	0
42	3	7	0,022	A	A	1	A
43	1	9	0,024	G	E	0	0
44	1	2	0,025	G	D	0	0
45	1	3	0,038	G	A	0	0
46	1	44	0,144	G	C	0	0

A imagem do banco de imagens disponibilizadas por Zooweb Karyotypes apresenta o cariótipo em pares e em grupos. Os número que estão na parte superior da imagem significam a ordem de geração das sub-imagens (rótulos) pelo software desenvolvido. Os números na parte inferior representam a ordem dos cromossomos e as letras de A até G significam os grupos. Como é observado nas figuras 35, 36, 37 e 38 têm-se os agrupamentos dos cromossomos por aproximação de centroide e vizinhança. Após observação dos casos foi percebido que quanto maior o coeficiente de aproximação mais erros pode-se ter, verificando este padrão nas classificações nota-se que quando o coeficiente > 0,6 têm-se mais erro. Aplicando a regra a imagem de análise o reconhecimento fica em 80%.






36 e 24	14 e 3	7 e 18	47 e 41	26 e 34
				
1 e 2	3 e 4	5 e 6	7 e 8	9 e 10
GRUPO A			GRUPO B	

FIGURA 35 – CROMOSSOMOS AGRUPADOS – GRUPO A, B

10 e 23	39 e 28	38 e 13	29 e 30	8 e 32	22 e 45	31 e 16
						
11 e 12	13 e 14	15 e 16	17 e 18	19 e 20	21 e 22	23 e 24
GRUPO C						

FIGURA 36 – CROMOSSOMOS AGRUPADOS – GRUPO C







6 e 40	2 e 25	11 e 12	27 e 43	42 e 46	21 e 9
					
25 e 26	27 e 28	29 e 30	31 e 32	33 e 34	35 e 36
GRUPO D			GRUPO E		

FIGURA 37 – CROMOSSOMOS AGRUPADOS – GRUPOS D, E







19 e 17	33 e 15	20 e 35	1 e 37	44 e 4	5
					
37 e 38	39 e 40	41 e 42	43 e 44	45 e 46	47
GRUPO F		GRUPO G		GRUPO C	GRUPO G

FIGURA 38 – CROMOSSOMOS AGRUPADOS – GRUPOS F,G,C,G

4.3.2 CÁLCULO DE CARACTERÍSTICAS DO HISTOGRAMA LBP

A primeira etapa é o processamento das sub-imagens para a criação de uma imagem padrão LBP, o próximo passo é a criação de um histograma para cada imagem.

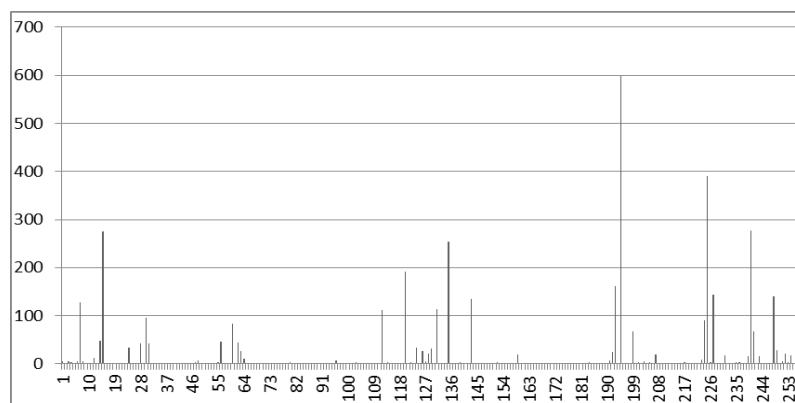


FIGURA 39 - HISTOGRAMA LBP SUB-IMAGEM 3

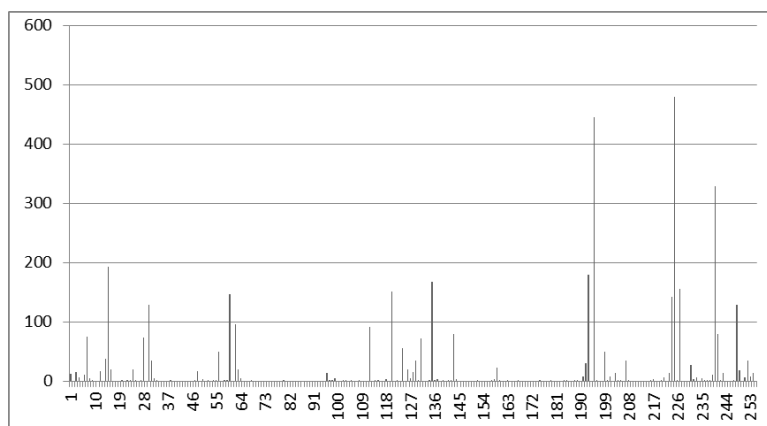


FIGURA 40 - HISTOGRAMA LBP SUB-IMAGEM 24

No cálculo de característica do histograma LBP faz-se uma comparação de similaridade do histograma da imagem LBP gerada utilizando o método estatístico CHI-QUADRADO. Como pode-se observar, na tabela a seguir o coeficiente de aproximação foi de 348,281, observando visualmente o histograma podemos observar que estes são muito similares.

TABELA 11 – RESULTADO LBP

LBP							
	C1	C2	DISTÂNCIA	GRUPO C1	GRUPO C2	VERIFICAÇÃO	ACERTO
3	3	24	348,281	A	A	1	A
7	7	24	467,026	A	A	1	A
14	14	24	289,153	A	A	1	A
18	18	36	294,585	A	A	1	A
24	24	14	289,153	A	A	1	A
36	36	18	294,585	A	A	1	A

A tabela 12 mostra o reconhecimento do método LBP, histograma e comparação chi quadrado. A taxa de reconhecimento fica em 55%.

TABELA 12 – VALIDAÇÃO LBP

LBP							
	C1	C2	DISTÂNCIA	GRUPO C1	GRUPO C2	VERIFICAÇÃO	ACERTO
1	1	33	227,223	G	F	0	0
2	2	12	257,472	D	D	1	D
3	3	24	348,281	A	A	1	A
4	4	34	355,987	C	B	0	0
5	5	17	182,459	G	F	0	0
6	6	9	229,557	D	E	0	0
7	7	24	467,026	A	A	1	A
8	8	11	193,533	C	D	0	0
9	9	27	204,072	E	E	1	E
10	10	23	163,066	C	C	1	C
11	11	31	180,560	D	C	0	0
12	12	25	137,721	D	D	1	D
13	13	28	235,804	C	C	1	C
14	14	24	289,153	A	A	1	A
15	15	17	209,224	F	F	1	F
16	16	23	195,464	C	C	1	C

LBP							
	C1	C2	DISTÂNCIA	GRUPO C1	GRUPO C2	VERIFICAÇÃO	ACERTO
17	17	5	182,459	F	G	0	0
18	18	36	294,585	A	A	1	A
19	19	35	144,618	F	G	0	0
20	20	43	389,244	G	E	0	0
21	21	17	198,747	E	F	0	0
22	22	29	249,456	C	C	1	C
23	23	10	163,066	C	C	1	C
24	24	14	289,153	A	A	1	A
25	25	12	137,721	D	D	1	D
26	26	10	267,472	B	C	0	0
27	27	9	204,072	E	E	1	E
28	28	29	211,112	C	C	1	C
29	29	28	211,112	C	C	1	C
30	30	31	171,289	C	C	1	C
31	31	30	171,289	C	C	1	C
32	32	6	255,294	C	D	0	0
33	33	11	202,017	F	D	0	0
34	34	18	299,144	B	A	0	0
35	35	19	144,618	G	F	0	0
36	36	18	294,585	A	A	1	A
37	37	35	180,277	G	G	1	G
38	38	28	222,471	C	C	1	C
39	39	8	223,832	C	C	1	C
40	40	42	523,747	D	E	0	0
41	41	40	1071,800	B	D	0	0
42	42	40	523,747	E	D	0	0
43	43	20	389,244	E	G	0	0
44	44	45	204,079	C	C	1	C
45	45	44	204,079	C	C	1	C
46	46	47	1356,200	E	B	0	0
46	47	46	1356,200	B	E	0	0

4.4 ESCOLHA DO MÉTODO

Com a aplicação dos Métodos descritos (Item 3.5) e a análise dos resultados (Item 4) conclui-se que o Método 3 apresentou-se mais adequado para a continuidade do projeto, apresentando respostas mais consistentes no seu desempenho. Os métodos 1 e 2 apresentaram resultados insatisfatórios e não serão adotados como projetos futuros.

Os resultados obtidos com o projeto exibidos neste capítulo servem como base para uma análise comparativa dos métodos, realizada no capítulo 5. Discussão.

5 DISCUSSÃO

A diferença entre os Métodos em comparação ao Método genérico para o processamento de imagens está evidenciado na Figura 41, que exhibe as etapas Aquisição, Pré-processamento, Extração das Características, Reconhecimento e Interpretação e Resultado.

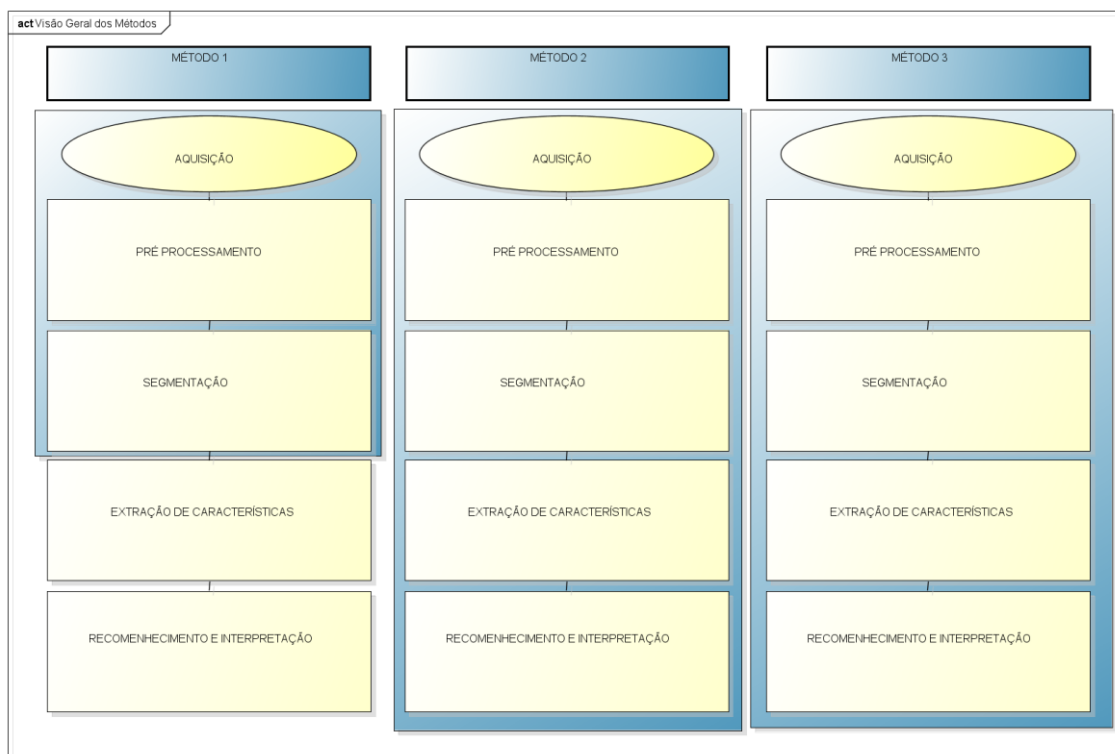


FIGURA 41 - VISÃO GERAL DOS MÉTODOS

A Figura 41 mostra uma Visão Geral dos Métodos exibindo a diferença comparativa sobre o Método genérico de Processamento de Imagem Digital (PDI).

Observando a figura 41 o Método 1 sobre o Método genérico para o processamento de imagens, tem-se que o processamento foi até a etapa de segmentação. O processamento não pôde prosseguir para as outras etapas, devido a baixa qualidade das imagens, baixa qualidade em questões de resolução, nitidez e ruídos por toda a imagem. O ruído presente em toda a imagem foi o elemento principal para a ineficácia do método, sendo que a aquisição da imagem é um fator determinante para o sucesso do PDI. Na etapa

de melhoria do contraste da imagem, consegue-se verificar que é realçada a imagem e evidenciado o ruído. No processo de limiarização existe a perda de informação da imagem, observando na imagem que após o processo algumas regiões importantes dos cromossomos são suprimidas. Devido a perda de informações na imagem as bordas, encontradas pelo algoritmo de detecção de bordas, foram dos cromossomos subdivididos. O processo de segmentação dos rótulos tem como referência os contornos encontrados na imagem, como em alguns casos os contornos não são bem definidos não se obteve bons resultados. Os resultados insatisfatórios não permitiram a progressão nos processos de extração de características de reconhecimento/interpretação.

O Método 2 foi até a etapa de reconhecimento e interpretação do Método genérico para o processamento de imagens, mas os resultados não foram satisfatórios. Houve a preparação das imagens através de processo manual, onde os melhores cromossomos do par 1 foram selecionados para testes da metodologia. Posteriormente a utilização de parte da Metodologia 1 para a geração das formas (bordas) dos cromossomos. As formas foram submetidas ao algoritmo para a parametrização PCA onde foi gerado um modelo a partir destas formas. A partir destes modelos (Shape Model) a forma criada buscava a qual outra forma ela era similar, esta varredura era feita sobre a imagem contendo todos os cromossomos. Modificações estavam sendo executadas no algoritmo, mas não estavam gerando resultados satisfatórios. Algumas etapas que foram realizadas manualmente no pré-processamento como a segmentação das bordas do cromossomo 1, teria de ser implementada no software, assim como mais característica e ajustes deveriam ser feitos para que o algoritmo obtivesse uma melhor eficácia.

O Método 3 iniciou a etapa de Pré processamento aplicando a escala de 200%, com o objetivo de obter mais informações após a aquisição da imagens (Figura 41). A ampliação da imagem foi de extrema importância as etapas posteriores. Algumas técnicas de limiarização foram testadas, mas para este caso o algoritmo de LAT foi o que obteve um melhor desempenho em sua função, em casos pontuais alguns cromossomos foram divididos ao meio, mas por consequência da qualidade das imagens. A etapa de morfologia matemática, na grande maioria dos casos preencheu perfeitamente os orifícios deixados pela

Limiarização, exceto em alguns casos pontuais citados anteriormente. A etapa de rotulação foi de extrema importância, pois através dela foi possível manipular e medir cada sub-imagem separadamente. O cálculo da altura do esqueleto, ou seja, altura do cromossomo foi eficaz em casos onde o cromossomo apresentava-se alinhado e curvado, mas em alguns casos quando partes deles ficavam encostadas umas nas outras o algoritmo não calculou corretamente sua altura. A transformação da imagem LBP foi uma tentativa de caracterização dos cromossomos através da textura, acreditava-se que as bandas dos cromossomos neste caso seriam determinantes na geração dos grupos. As características de cada Método, suas subdivisões, tipo de amostra e detalhes de PDI estão evidenciados na tabela 13.

TABELA 13 – CARACTERÍSTICAS DOS MÉTODOS 1, 2 E 3.

ETAPAS PDI	MÉTODO 1	METODO 2	METODO 3/1	METODO3/2
Amostra	Akodon montensis	Akodon montensis e Homo sapiens	Homo sapiens	Homo sapiens
Pré processamento	Equalização de Sub-imagens	Escolha visual dos melhores cromossomos Grupo A	ESCALA Equalização de Sub-imagens	ESCALA Equalização de Sub-imagens
Segmentação	Limiar de Otsu Detecção de Bordas Convex Hull	Recorte manual dos melhores cromossomos do grupo A Detecção de Bordas	Limiar de LAT Abertura morfológica Binária Inversão da Imagem Rotulação Esqueletização Binária	Limiar de LAT Abertura morfológica Binária Inversão da Imagem Rotulação Imagem padrão LBP
Extração de Características	Não foi criado o algoritmo, pois a metodologia de rotulação não se mostrou eficiente. Isso impossibilitou a medição.	Medidas Geradas por parametrização PCA	Medidas de Área, Perímetro, Altura e Fator compacidade	Histograma da imagem LBP
Reconhecimento e interpretação	-	Comparação do modelo PCA com toda a imagem.	Análise de similaridades das características por meio do software IBM SPSS, Cluster Hierárquico > Centróide > distância euclidiana ao quadrado.	Análise-se as similaridades entre os histogramas por meio do método estatístico chi-quadrado.
Observações	-	-	O procedimento se mostra eficaz para o agrupamento dos cromossomos nas classes A,B,C,D,E,F e G, estes agrupamentos possuem medidas semelhantes.	O procedimento se mostra eficaz para o agrupamento dos cromossomos nas classes A e C estes agrupamentos possuem características LBP semelhantes, para os grupos B, D, E,F e G, o algoritmo não mostrou-se muito eficaz.
Resultados	-	Pouca similaridade com os cromossomos no conjunto de reconhecimento	A média de acerto: Somente área: 70% Todas as Medidas 65%	A média de acerto: Histograma LBP: 56%

A substituição do banco de imagens, com a troca dos tipos de cromossomos de *Akodon montensis* para *Homo sapiens*, pôde ser implementada porque não interferiu no projeto inicial de Segmentação e Análise de Imagens digitais de Metáfases Cromossômicas e o mesmo método genérico foi mantido.

O Método 1, utilizando o resultado do limiar como máscara, conseguiu separar o objeto de interesse do fundo da imagem, porém como o ruído está por toda a imagem, algumas informações foram perdidas. Devido a baixa qualidade das imagens o algoritmo criou inúmeros rótulos com ruídos. A implementação ficou complexa devido aos tipos de estruturas utilizadas no OpenCV e a dificuldade na conversão de tipos entre elas, não sendo eficaz testes com o novo banco de imagens.

A predeterminação da nova base de imagens do banco Zooweb Karyotypes (2013), foi justificada com a modificação do Método 1, a complexidade elevada poderia prejudicar o desempenho do algoritmo. O Método 2 os resultados encontraram similaridades entre os cromossomos, o algoritmo precisaria sofrer modificações para encontrar o modelo correto, e outras etapas precisariam ser implementadas. A complexidade do banco de imagens do Método 1, as adaptações e novas implementações do Método 2, justificaram a escolha do Método 3.

No Método 3 as imagens não possuem um padrão definido para aquisição ou para número de cromossomos, possuem baixa resolução, pouca nitidez; o que dificulta o processamento de imagens. A utilização do método escala em 200%, baseado na Pirâmide Gaussiana (BRADSKY; PISAREVSKY; BOUGUET, 2006), foi suficiente para executar o aperfeiçoamento da quantidade de informações presentes na imagem. A limiarização utilizando o algoritmo LAT foi utilizada porque o algoritmo de Otsu utilizado no Método 1 não foi eficaz para a nova base de imagens. O método de rotulação utilizado na Metodologia 1 poderia ser utilizado na Metodologia 3, mas a implementação da rotulação do método 3 foi a que apresentou melhores resultados, assim não necessitando de muitas intervenções que deveriam ser realizadas na rotulação da Metodologia 1.

Nos resultados no Método 3 houve aumento da similaridade entre os pares de cromossomos, o vetor de características LBP trouxe uma maior

confiabilidade ao resultado final, é possível separar os cromossomos em seus devidos grupos, A, B, C, D, E, F, G, com poucos erros. Estes grupos são formados de acordo com as características de área, perímetro e altura. Após a implementação do LBP deseja-se que os cromossomos sejam identificados com seus devidos pares.

Neste projeto inicialmente foram utilizadas imagens de cromossomos do roedor *Akodon montensis* e de cromossomos humanos (*Homo sapiens*), futuramente serão adquiridas novas imagens do roedor *Akodon montensis* e o software poderá ser adaptado para esta nova base de imagens.

Com a realização deste estudo, percebe-se que a abordagem multiprofissional e interdisciplinar são fundamentais para o desenvolvimento, visto que uma série de conceitos de processamento de imagens auxilia a análise dos cromossomos.

Nos resultados do Método 3 houve aumento da similaridade entre os pares de cromossomos, o vetor de características LBP trouxe uma maior confiabilidade ao resultado final, é possível separar os cromossomos em seus devidos grupos, A, B, C, D, E, F, G, com poucos erros. Estes grupos são formados de acordo com as características de área, perímetro e altura. Após a implementação do LBP os cromossomos são identificados com seus devidos pares.

No método 3/1 a análise de medidas de área obteve os melhores resultados, exibindo a validação de todos os resultados com 76% de acerto, sendo somente área 70%, e em todas as medidas 65%. O método 3/2 obteve o reconhecimento do método LBP, histograma e comparação chi quadrado, com taxa de reconhecimento em 55%. Os resultados do histograma LBP trouxeram um nível de reconhecimento aceitável, porém a caracterização por medidas teve maior reconhecimento.

6 CONCLUSÃO

O objetivo principal deste projeto, construir uma ferramenta computacional para facilitar o processo de identificação de cromossomos humanos, foi atingido. Os objetivos específicos de segmentar e analisar imagens digitais de cromossomos metafásicos tratados com a técnica de bandeamento G, através dos métodos de processamento de imagens; e analisar o desempenho dos resultados nos métodos de processamento de imagens; também foram alcançados.

Após diversos testes e estudos verificamos que o método 3, nas suas duas etapas, teve desempenho e resultados esperados, no agrupamentos das classes cromossômicas A, B, C, D, E, F, G. Uma etapa fundamental para estudos futuros é a criação de um protocolo de aquisição das imagens através de um microscópio. Desenvolvimentos de novas caracterizações dos cromossomos aumentarão o reconhecimento, e posteriormente facilitará na tarefa de especiação pelo geneticista.

Na análise dos resultados obtidos observamos que o sistema obteve no método 3/1, com a análise de medidas de área, os melhores resultados, exibindo a validação de todos os resultados de 76% de acerto, dentre os testes realizados. O método 3/2 obteve o reconhecimento do método LBP, histograma e comparação chi quadrado, com taxa de reconhecimento em 55%. Os resultados do histograma LBP trouxeram um nível de reconhecimento aceitável, porém a caracterização por medidas teve maior reconhecimento. As análises por meio do software IBM SPSS, mostrou qual metodologia é mais eficaz para o reconhecimento dos grupos. Futuramente estas metodologias poderão ser implementadas no software.

As imagens adquiridas do acervo fotomicrográfico da UFPR, mostraram limitações quanto ao processamento devido sua baixa qualidade. Neste projeto três metodologias foram construídas, sendo que os experimentos iniciais nos métodos 1 e 2 não foram tão eficazes quanto o 3. O comparativo entre as metodologias mostram isso. Dependendo do banco de imagens as metodologias 1 e 2 poderiam ter sido continuadas, mas devido a complexidade deste conjunto o projeto prosseguiu somente no método 3. Algumas limitações da pesquisa

foram observadas ao longo do projeto, primeiramente a utilização de um banco de imagens de baixa qualidade. Dificultando o PDI e a busca de uma metodologia que tivesse uma melhor eficácia. A pouca disponibilidade de bancos de imagens cromossômicas na internet. O banco encontrado disponibilizava poucas imagens, com baixa resolução. Problemas de qualidade na imagem acarretam em medidas nem sempre precisas, o que diminui o acerto, pois as variações de tamanho entre as classes nem sempre são grandes. A criação de uma metodologia eficaz para o PDI foi de extrema importância, pois para projetos futuros outras técnicas de PDI e características novas poderão ser agregadas ao software desenvolvido. Algumas análises em outros softwares como o SPSS mostram que estas características podem compor o software desenvolvido futuramente. Funcionalidades novas podem ser desenvolvidas a fim de classificar os cromossomos em ordem, par e em análises de padrões de espécies para espécies. Assim chegamos à conclusão que o sistema desenvolvido em sua primeira versão atende aos requisitos propostos neste projeto.

7 REFERÊNCIAS

- 1) ANTUNES, A. F. B. Classificação de Ambiente Ciliar Baseada em Orientação a Objeto em Imagens de Alta Resolução Espacial. Tese apresentada ao Curso de Pós-Graduação em Ciências Geodésicas, Setor de Ciências da Terra. Universidade Federal do Paraná. Curitiba, 2003.
- 2) BARROS, Beatriz Amstalden et al . A inclusão de novas técnicas de análise citogenética aperfeiçoou o diagnóstico cromossômico da síndrome de Turner. Arq Bras Endocrinol Metab, São Paulo, v. 53, n. 9, Dec. 2009 . Available from <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0004-27302009000900010&lng=en&nrm=iso>. Acesso em 17 de abril 2013. <http://dx.doi.org/10.1590/S0004-27302009000900010>.
- 3) BERTOLOTTO, C. E. V. Enyalius (Leiosauridae, Squamata): O que os dados moleculares e cromossômicos revelam sobre esse genero de lagartos endêmicos do Brasil. Tese (Doutorado) - Instituto Biociência, Universidade de São Paulo. São Paulo, SP. (2006).
- 4) BORGES - OSÓRIO, M.R.; ROBINSON, W.M. Genética Humana. Editora Artmed, 2ª edição, Porto Alegre, 2002.
- 5) BRADSKY, Gary; KAEHLER, Adrian. Learning OpenCV. Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472. ISBN: 978-0-596-51613-0. September 2008.
- 6) BRUNONI, Décio. Aconselhamento Genético. Ciênc. saúde coletiva, São Paulo, v. 7, n. 1, 2002 . Available from <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-81232002000100009&lng=en&nrm=iso>. Acesso em 11 de abril 2013. <http://dx.doi.org/10.1590/S1413-81232002000100009>.
- 7) CASPERSON, T; ZENCH, L; JOHANSSON, C. Differential banding of alkylating fluorochromes in human chromosomes. Exp. Cell Res. 1970; 60:315-19.
- 8) DAUBECHIES, I. The wavelet transform, time-frequency localization and signal analysis. IEEE Transactions on Teoria da Informação.

Setembro 1990. Departamento, Michigan Univ., Ann Arbor, MI, EUA. Volume: 36, Issue: 5. p (s): 961-1005.

9) De Berg, van Kreveld, Overmars, Schwarzkopf. Computational Geometry: Algorithms and Applications. 2nd edition, Springer-Verlag

10) DeCS. Descritores em Ciências da Saúde. Biblioteca Virtual em Saúde. Disponível em: <http://decs.bvs.br/>. Acesso em 12/04/2013.

11) DLUGOSZ, F. L.; ROSOT, N. C.; ROSOT, M. A. D.; OLIVEIRA, Y. M. M. ÍNDICE PARA A AVALIAÇÃO DE SEGMENTAÇÃO DE IMAGENS. Depto. de Ciências Florestais. EMBRAPA Florestas, Colombo. Doutorando em Engenharia Florestal, UFPR. FLORESTA, Curitiba, PR, v. 39, n. 1, p. 131-143, jan./mar. 2009.

12) DUTRILLAUX, B. & COUTURIER, J. 1981. La Pratique de l'Analyse Chromosomique. - Masson, Paris.

13) FACON, J. Processamento de Imagens. Pontifícia Universidade Católica do Paraná. Curso Mestrado em Informática Aplicada. 2002.

14) FELIPE, Rafaela Dandolini; NEVES, L. A. P. Pre-printed and Hand-filled Table-Form Analysis aiming Cell Extraction. In: The Eighth IAPR International Workshop on Document Analysis Systems (DAS 2008), 2008, Nara. The Eighth IAPR International Workshop on Document Analysis Systems. Nara: IEEE Computer Society, 2008. v. 1. p. 439-443.

15) FENG, Xinwei; CONG, Peisheng; ZHU, Zhongliang; DU, Xinyuan. Automated pairing of human chromosomes applying gradient profile and similarity matching algorithm. Chemometrics and Intelligent Laboratory Systems. 46–52. Elsevier B.V. doi:10.1016/j.chemolab. 2011.

16) GARCIA, Eloi S.; CHAMAS, Claudia Inês. Genética molecular: avanços e problemas. Cad. Saúde Pública, Rio de Janeiro, v. 12, n. 1, Mar. 1996. Disponível em http://www.scielo.org/scielo.php?script=sci_arttext&pid=S0102-311X1996000100022&lng=en&nrm=iso. Acesso em 11 de abril de 2013. <http://dx.doi.org/10.1590/S0102-311X1996000100022>.

17) GONZALES, R. C; WINTZ, P. Digital image processing. Reading: Addison Wesley, 1987. 431 p.

- 18) GONZALES, R. C; WOODS, R. Processamento de imagens digitais. São Paulo: E. Blücher, 2002. 508 p.
- 19) GONZALEZ, R. C.; WOODS, R. E. Digital image processing. Reading, MA: Addison-Wesley, 1992. 716p.
- 20) GOSE, Earl, JOHNSONBAUGH, Richard, JOST, Steve PARTNER RECOGNITION AND IMAGE ANALYSIS, 1a Edição. New Jersey, Prentice Hall PTR, 1996, pp 329-335.
- 21) GUERRA, M. Introdução a citogenética geral. Rio de Janeiro: Ed. Guanabara, 1988, 142p.
- 22) J. Canny, A computational approach to edge detection, IEEE Transactions on pattern analysis and machine intelligence, 8(6), pp 679 – 698 (1986).
- 23) J. Facon, “Morfologia Matemática: Teoria e Exemplos”, Editor Jacques Facon, Curitiba, 1996
- 24) Karp, Gerald. Cell and Molecular Biology: Concepts and Experiments . 5ª ed. New Jersey: John Wiley, 2008. p. 390-395.
- 25) KASAHARA, S. (2001). Práticas de Citogenética. Universidade Estadual Paulista, Rio Claro. 69p
- 26) KURTZ, G. C; BONINI, T.; LIBRELOTTO, G. R.; PERLES, L. A.; SAGRILLO, Michele. Identificação Automática de Cromossomos. Anais do VII Simpósio de Informática da Região Centro do RG-SIRC/RS - ISBN 978-85-88667-89-1. Santa Maria - RS, junho 2008.
- 27) KURTZ, Guilherme Chagas. METODOLOGIAS PARA DETECÇÃO DO CENTRÔMERO NO PROCESSO DE IDENTIFICAÇÃO DE CROMOSSOMOS. UNIVERSIDADE FEDERAL DE SANTA MARIA. CENTRO DE TECNOLOGIA PROGRAMA DE PÓS - GRADUAÇÃO EM INFORMÁTICA METODOLOGIAS PARA DETECÇÃO DO CENTRÔMERO NO PROCESSO DE IDENTIFICAÇÃO DE CROMOSSOMOS. DISSERTAÇÃO DE MESTRADO. Santa Maria, RS, Brasil. 2011.
- 28) L.Li, J. Gong, W. Chen, “Gray-level Image thresholding based on Fisher linear projection of two-dimensional histogram”, Research Institute of Automation, Southeast University, Nanjing 210018, People’s Republic of China , 1996.

29) MARQUES FILHO; VIEIRA NETO, H. Processamento Digital de Imagens, Rio de Janeiro: Brasport, 1999. ISBN 8574520098.

30) MARTINS, A. de M. Contribuições aos Processos de Clustering com Base em Métricas não-Euclidianas / Allan de Medeiros Martins. Dissertação (Mestrado) - Universidade Federal do Rio Grande do Norte. Centro de Tecnologia. Programa de Pós-Graduação em Engenharia Elétrica. Natal: 2005.

31) MARTINS, Ana Luiza B.; Barbosa-Gonçalves Adriana; Rocha, Simone S.; Conte, Agnes Cristina Fett. Análise citogenética por bandamento GTG convencional e em alta resolução da região em pacientes com doenças do espectro autístico. Arq Ciênc Saúde, 2009. p. 115. Disponível em: <http://www.cienciasdasaude.famerp.br/racs_ol/vol-16-3/IDJ3.pdf>. Acesso em: 12/04/2013.

32) NEVES, S. C. M.; PELAES, E. G. ESTUDO E IMPLEMENTAÇÃO DE TÉCNICAS DE SEGMENTAÇÃO DE IMAGENS. Revista Virtual de Iniciação Acadêmica da UFPA. Universidade Federal do Pará – Departamento de Engenharia Elétrica e de Computação – Laboratório de Processamento de Sinais (UFPA – DEEC – LaPS). Campus Universitário do Guamá – Belém – Pará. Vol 1, No 2, Julho 2001. Pag. 1-11. Disponível em: <http://www.ufpa.br/revistaic>. Acesso em: 23/04/2013.

33) OLIVEIRA, J. de A.; DUTRA, L. V.; RENNÓ, C. D.; SANTOS, P. S. Extração de Atributos de Forma para Classificação de Imagens de Alta Resolução do Satélite. HRC/CBERS-2B. Universidade Federal do Recôncavo da Bahia – BA, Brasil. 2008. Disponível em: <http://mar.te.dpi.inpe.br/col/dpi.inpe.br/sbsr@80/2008/11.17.22.33/doc/7015-7022.pdf>. Acesso em: 15/05/2013.

34) OTSU, N, "A threshold selection method from gray-level histograms," IEEE Transactions on Systems, vol. 9, pp. 62-66, 1979.

35) PECCININI-SEALE D; CARABETTA JR, W. & GREGORIN, S.A. 1977a; Heterocromatina constitutiva em cromossomos de lacertílios brasileiros. Ciênc. Cult. Suppl. 29 (7): 709.

36) PECCININI-SEALE, D.; GREGORIN, S.A.; CARABETTA JR, W. 1977b. Bandamento em cromossomos de lacertílios brasileiros tratados com tripsina. Ciênc. Cult. Suppl. 29 (7): 709.

- 37) Queiroz, José Eustáquio Rangel; Gomes, Herman Martins. Introdução ao Processamento Digital de Imagens. Revista RITA. Volume VIII. Número 1. 2001. Disponível em: <<http://www.dsc.ufcg.edu.br/~hmg/disciplinas/graduacao/vc-2011.2/Rita-Tutorial-PDI.pdf>>. Acesso em 15/05/2013.
- 38) REGATEIRO, F. J. Manuel de Genética. Imprensa Universidade de Coimbra. Coimbra, 2007. ISBN 972-8704-12-7
- 39) Sanqiang Zhao, Yongsheng Gao, Baochang Zhang. Sobel-LBP. In Proceedings of the International Conference on Image Processing, ICIP 2008, October 12-15, 2008, San Diego, California, USA. pages 2144-2147, IEEE, 2008. [doi]
- 40) SEABRIHT, m. 1971. A rapid banding technique for human chromosomes. Lancet. 1971: 971-972.
- 41) SEIXAS, F.L.; MARTINS, A.; STILBEN, A. R.; MADEIRA, D.; ASSUMPÇÃO, R.; MANSUR, S.; VICTER, S. M.; MENDES, V. B.; CONCI, A. AVALIAÇÃO DOS MÉTODOS PARA A SEGMENTAÇÃO AUTOMÁTICA DOS TECIDOS DO ENCÉFALO EM RESSONÂNCIA MAGNÉTICA. Universidade Federal Fluminense. Rua Passo da Pátria, 156. SPOLM 2008 ISSN 1806-3632 Rio de Janeiro, Brasil, 2008.
- 42) Shapiro, L., and Stockman, G. (2002). Computer Vision. Prentice Hall. pp. 69–73.
- 43) Somasundaran, D. Nirmala, M. 2010. Automatic Segmentation And Karyotyping Of Chromosomes Using Bio-Metrics.
- 44) SPRING – Sistema de processamento de informações georreferenciadas. Processamento de imagens. São José dos Campos: NETGIS, 1998. v.2.
- 45) STEDMAN, Thomas L. STEDMAN'S MEDICAL DICTIONARY, 25th edition. 1995
- 46) SUMNER, A.T. 1972. A simple technique for mostraring centromeric heterochromatin. Experimental Cell Research 75: 304-306.
- 47) Toru Abe, Chieko Hamada, Tetsuo Kinoshita: A Chromosome Image Recognition Method Based on Subregions. ACCV (3) 2009: 611-620

- 48) VALENZUELA, C; ANTELO, R. D.; ADOLFO, T. Aspectos Bioéticos y Legales de la Investigación Genética en Miembros Menores de Edad de una Familia con Adrenoleucodistrofia Ligada al X de Tucumán, Argentina. *rev.latinoam.bioet.*, Bogotá, v. 10, n. 1, June 2010. Available from <http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1657-47022010000100009&lng=en&nrm=iso>. Acesso em 11 de abril de 2013.
- 49) WANG , Yum, Image Enhancement Based on Equal Area Dualistic Sub-Image Histogram Equalization Method, *IEEE Transaction on Consumer Eletronics*, Vol 45, No 1Feb, 1999, 68-75.
- 50) WANG, Yu; CHEN, Qian; ZHANG, Baomin. "Image enhancement based on equal area dualistic sub-image histogram equalization method," *Consumer Electronics, IEEE Transactions on* , vol.45, no.1, pp.68,75, Feb 1999. doi: 10.1109/30.754419.
- 51) WOODS, Richard; GONZALEZ, Rafael C. *Processamento de Imagens Digitais*. Editora Edgard Blucher. ISBN: 9788521202646. 2000.
- 52) Z. Xiong, Q. Wu, and K. Castleman, Enhancement, classification and compression of chromosome images. , " *Proc. Workshop on Genomic Signal Processing and Statistics (GENSIPS)*, Raleigh, NC, October 2002.
- 53) ZECH, L. 1973. Fluorescence banding techniques. *Nobel Symp.* 23: 28-31.
- 54) Zhang, T. Y.; Suen, C. Y. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, v. 27, n. 3, p. 236-239, 1984.
- 55) Zooweb Karyotypes. Human Karyotypes for teaching. Wisconsin State Laboratory of Hygiene. Disponível em: <http://worms.zoology.wisc.edu/zooweb/Phelps/karyotype.html>. Acesso em: 17/04/2013.